Aalto University
School of Science
Degree Programme of Computer Science and Engineering

Heikki Lindroos

Effects of Social Presence on the Viewing Experience in a Second Screen Environment

Master's Thesis
Espoo, January 27, 2014

Supervisor: Professor Pirkko Oittinen, Aalto University
Instructor: Ph.D. (tech) Teemu Kinnunen

Aalto University
School of Science
Degree Programme of Computer Science and Engineering

ABSTRACT OF
MASTER'S THESIS

Author: Heikki Lindroos

Title:
Effects of Social Presence on the Viewing Experience in a Second Screen Environment

Date: January 27, 2014 **Pages:** viii + 97

Professorship: Mediatechnology **Code:** T-75

Supervisor: Professor Pirkko Oittinen

Instructor: Ph.D. (tech) Teemu Kinnunen

Technological progress and changes in society have affected the media consumption habits that people are now accustomed to. The introduction of smartphones, the penetration of social networking services and the increased presence of both have put the consumer in a more active role. In the context of watching TV, these changes have brought up new surroundings for viewing media. This is known as a second screen viewing environment, in which the TV viewer multitasks with another electronic device simultaneously. Second screen content can vary widely, here we are interested in the social media content.

Both broadcasting companies and the research community have shown interest towards the field. Many have selected Twitter as the platform of choice to accompany viewers. Tweets are used both among consumers and between consumers and content providers. Researchers have studied the new multitasking environment from a traditional performance and gratification perspective.

In this thesis the effects of second screen viewing are explored on the viewing experience. Differences are sought between two viewing conditions, one with plain TV viewing and one where TV viewing is accompanied by a tweet feed on the second screen. In addition, the differences in effect between TV genres are studied.

As a result it was found that even though the two media continue to merge, people still differentiate them. Further, it can be reported that significant differences in the suitability of second screen content between the genres were found. Moreover, identifiable factors affecting second screen content amenity are reported.

Keywords: social media, Twitter, television, viewing experience

Language: English

Aalto-yliopisto
Perustieteiden korkeakoulu
Tietotekniikan tutkinto-ohjelma

DIPLOMITYÖN TIIVISTELMÄ

Tekijä: Heikki Lindroos

Työn nimi:
Tabletilla tapahtuvan sosiaalisen läsnäolon vaikutus television katselukokemukseen

Päiväys:	27. tammikuuta 2014	**Sivumäärä:**	viii + 97
Professuuri:	Mediatekniikka	**Koodi:**	T-75
Valvoja:	Professori Pirkko Oittinen		
Ohjaaja:	TkT Teemu Kinnunen		

Uudet innovaatiot, palvelut ja tuotteet mullistavat jatkuvasti mediateollisuutta. Uuden sukupolven mobiililaitteet ja sosiaalisen median voimakas läpilyönti ovat asettaneet television katselijan aikaisempaa aktiivisempaan asemaan. Tämä muutos on samalla ajanut kuluttajat seuraamaan useampaa mediaa ja laitetta samanaikaisesti. Television katselua tukevaa toista laitetta on nimitetty englanninkielisessä kirjallisuudessa termillä *second screen*.

Sekä median tuottajat että tutkijayhteisö ovat kiinnittäneet huomiota tähän kehitykseen. Sisällön tuottajat ovat valinneet sosiaalisen median alustaksi Twitterin, palvelun, jossa käyttäjät voivat jakaa lyhyitä, korkeintaan 140 merkin mittaisia viestejä mm. koskien katsomiaan televisio-ohjelmia. Tutkijat ovat keskittyneet tutkimaan tätä uutta ympäristöä mm. etsimällä kasvavan mediakäytön tuomaa mielihyvää ja katselutottumuksien seuraamuksia. Moni tutkimus sijoittuu myös hyvin väkirikkaisiin maihin, joissa mm. Twitterin käyttö on huomattavasti yleisempää kuin Suomessa.

Diplomityössä on tutkittu tätä ilmiötä suomenkielisessä ympäristössä. Työssä on selvitetty, kuinka ihmiset kokevat lisälaitteen tuoman hyödyn ja kuinka hyvin lisälaitteen tuoma sisältö sopii median eri tyylilajeihin. Työssä myös raportoidaan, miten ihmiset kokevat tämän uuden mediaympäristön.

Työn tuloksena onnistuttiin selvittämään, että medioiden integroitumisesta huolimatta ei niitä katselukokemuksen aikana silti koeta yhdeksi. Lisäksi tyylilajien välillä pystyttiin havaitsemaan huomattavia lisäsisällön sopivuuden välisiä eroja. Tämän lisäksi selvisivät mielyttäneeseen lisäsisältökokemukseen vaikuttaneet syyt.

Asiasanat:	sosiaalinen media, Twitter, televisio, katselukokemus
Kieli:	englanti

Acknowledgements

I wish to thank my supervisor Professor Pirkko Oittinen for quickly organizing a thesis position form me at the department and my instructor Ph.D. Teemu Kinnunen for all the help and support he provided.

I wish to thank my parents for their help and support throughout my years in school.

I wish to thank Mr. Kenta Kusumoto for doing an amazing job when programming the test environment software used in the experimental part of this thesis.

I would like to thank the people who participated in the experimental part of this study.

I also wish to thank Viestintäalan tutkimusäätiö[1] for financially supporting this thesis.

Espoo, January 27, 2014

Heikki Lindroos

[1] `http://www.vkl.fi/vts/vts`

Abbreviations and Acronyms

API	Application Programming Interface
BAS	Behavioral Activation System
BIS	Behavioral Inhibition System
Companion App	An application designed to accompany second screen viewing
ECG	Electrocardiography (also EKG), interpretation of the electrical activity of the heart
EDA	Electrodermal activity, also known as Galvanic skin response
EMG	Electromyography, a technique for recording and evaluating the electrical activity produced by skeletal muscles
ETG	Eye Tracking Glasses, a solution for gaze tracking
Hashtag	A method used by Twitter to tag topics (e.g. ”#thesis”)
MPI	Multitasking Preference Inventory, a measure of polychronicity
OR	Orienting Response, immediate response to a change in its environment
PNS	Parasympathetic nervous system
Retweet	On Twitter, the act of forwarding (or reposting) a tweet
SAM	Self Assessment Manikin, a nonverbal method used for emotional selfreporting
SCL	Skin Conductance Level
SCR	Skin Conductance Response
SNS	Sympathetic nervous system
U&G	Uses and Gratifications
iTV	interactive television
YLE	Finnish Broadcasting Company

Contents

Chapter 1

Introduction

From its introduction, television has prompted off-screen social engagement. At first though, television was considered to be a one-way media that isolated people. Yet, it is becoming increasingly difficult to ignore the social aspects that are built around TV and social media networks. Now, TV can be considered the trigger of 21st century "water-cooler-effect".

Through the times technical development has changed the social scene. Today, handheld smart devices are a de facto. Tablet and mobile phones play an increasingly important role in our everyday lives, accompanying us almost everywhere. This can also be seen in TV watching habits. The latest number Rideout et al. [60] suggests that close to 70% of Americans use an additional device while watching TV. The same trend can be seen also in Finland Statistics of Finland [74].

In association with TV viewing, the industry has labeled these devices as "second screens". This development have led to an increasing interest in creating new business models and services for these new platforms, most notably the introduction of *companion applications* Evolumedia group [23]. This is a software designed to enhance the viewing experience by enriching the content and providing increased social presence and a possibility for a deeper insight in the topic Evolumedia group [24].

Regardless of the hype, the first generation of these applications have not been able to fully penetrate into the home audiences. Thus, content providers try to adapt and to develop the applications further in the effort of enriching the previously passive act of watching TV into a new more pleasing interactive media multitasking event.

The objective of this thesis is to find the answers to the following research questions by a experimental study: i) *What are the effects of social presence on the viewing experience?*, i.e. studying i.i) *How does the social second screen content affect the viewing experience?*and i.ii) *How do people appraise*

the new environment?

This thesis presents the findings of an experimental study on involving a TV-tablet second screen viewing session. Chapter 2 begins by laying out the theoretical dimensions of the research, and looks at how the second screen has been brought up in literature. Chapter 3 describes the methods regarding subjective testing and metrics used in the experimental phase. The chapter will then go on by describing the design and execution of the experimental phase. In Chapter 4 the most important findings are presented in regard to the research questions presented.

The findings are considered in the light of the research questions in Chapter 5, where the a set of recommendations are presented to take into consideration when developing these types of applications. Finally, conclusions are drawn in Chapter 6.

Chapter 2

Background

Some consider watching TV to be a passive activity. This might still be true for some situations or programs. Quite often though, people are not that deeply immersed in the program and this provokes other simultaneous actions such as reading the program guide, talking or increasingly the use of a secondary device (laptop, tablet or mobile phone) [60, 82]. The multitasking environment pushes the boundaries of human cognitive capabilities [49]. Multitasking environments have been reported to decrease the immersion and performance, yet media multitasking is attractive and rewarding for the viewer.

The last few years have meant a massive expansion for social networking sites. Social networking has penetrated to almost all parts of our everyday life [52]. The introduction of user-friendly and nifty handheld devices has now brought social media and TV viewing together, forming an integral viewing environment and media experience.

2.1 Multitasking

Multitasking is the practice of processing multiple tasks simultaneously. It has been researched widely. Largely the results show that multitasking has a negative effect on task performance [1, 33, 38, 73]. Frequently reported findings are decreased memory performance and lesser immersion in tasks at hand compared to performing a single task at a time. Researches also highlight the negative effect of decreased information intake [49].

In contrast, some [54] underscore the positive effects on productivity and the welcome breaks during difficult tasks. Multitasking also allows ideas to mature while focusing on an alternative task at hand. Nonetheless, multitasking is here to stay, as we juggle with multiple tasks at home, at

school and at work. An interesting conflict arises between the prevalence of second screen use and the frequently reported impacts on performance.

Even though multitasking preference is individual, most humans tend to do multiple tasks simultaneously [56]. In addition, humans' ability to multitask has adapted over the years as the society and technology around us have changed [8, 60]. Modern technology is built around the concept of multitasking: operating systems allow multiple simultaneous programs to be run, web browsers have implemented tabs to allow multitasking and now even the multimedia environment has seen changes that provoke multitasking [1].

At first, multitasking might sound as a simple concept to grasp, but as explined here, multitasking can be looked at from multiple perspectives and disciplines.

2.1.1 Human Ability to Multitask

Ravaja [57] presents the literature definition and successfully sums up essential findings regarding the properties of human attention. There, attention is defined as the allocation of limited mental resources to a specific stimulus. Further, attention can be divided into two groups, i.e. *selective attention* and *capacity theories*. Selective attention refers to the process of selecting certain stimuli for processing while ignoring others and capacity theories state that humans have a limited sized pool of attention to allocate resources from. In this thesis, the allocation of attention is defined by the media the participant chooses to follow.

Novak et al. [49] state that performing many simultaneous tasks requires more cognitive resources. The capacity theory is supported by the fact that multitasking gets increasingly challenging as the cumulative cognitive load, set by the tasks at hand, increases, i.e. as the cumulative task complexity exceeds the recourses available, human performance starts to suffer. As to multitasking capabilities, human performance can be increased i) by freeing cognitive capacity for task execution or ii) by decreasing the required cognitive load of a specific task.

Novak et al. [49] reported differences in recall performance of low and high recall messages across all memory measures. The results are in favour of the high relevance memory events. Free recall tasks have been reported to result in more errors during multitasking compared with single task execution.

Lang [34] introduces a theory on how cognition works. According to her, input is processed by three sub-processes: encoding, storage, and retrieval. *Encoding* refers to selection of information from a stream or me-

dia and then encoding it to a mental representation. *Storage* is the phase when the representation is stored in to the mind. Finally, *retrieval* is the re-activation of the mental representation. These tree processes work in parallel. It is also noteworthy that the overload of one sub-process may affect the others. Further, Novak et al. [49] point out that overloading the brain at the encoding phase is more harmful for retrieval tasks than in the retrieval phase.

Yerkes and Dodson [86] have created a model of how human performance is affected by arousal levels which is known as the *Yerkes–Dodson law*. This curvilinear (U-shaped) inverted model states that with both high and low arousal levels human performance is suboptimal and the most efficient performance levels are found with moderate levels of arousal. Arousal reflects the level of attentiveness and alertness. At low levels the performance is distracted by lack of alertness, at high levels the task is distracted by high level disturbing responses. The optimal performance point is characterized by just the right amount of alertness and relaxation as explained by Palladino [53].

2.1.2 Concurrent, Interleaved and Sequential Multitasking

Adler and Benbunan-Fich [1] state that the amount of temporal overlap can be used to categorize multitasking tasks in three ways: sequential, parallel and interleaved. Each stands for a certain degree of concurrency of the tasks.

According to Adler and Benbunan-Fich [1], *sequential* multitasking can be argued not to be multitasking at all as the next task starts once the previous is finished. This does not fall in the category of traditional multitasking since there is actually no multitasking as no temporal overlap occurs. However, it can be used as a baseline condition for comparison with other types of multitasking.

In *interleaved* multitasking, the task at hand is voluntarily or involuntarily interrupted to allow the execution of another task. Later the original task is resumed, again voluntarily or involuntarily. An example of this type of multitasking is simultaneously making food and attending to a child.

Finally, during *parallel* multitasking all tasks are handled at the same time. This way there is a maximum amount of concurrency. This is often hard to achieve, unless different types of attention are required, e.g. listening to music and reading.

One can consider second screening to fall in to the parallel category with some characteristics of interleaved multitasking. As two or more types of media are simultaneously consumed, interruptive events happen on both screens.

2.1.3 Task Switching

Task switching is a well studied phenomenon. Multitasking performance is closely dependent on the cost of switching between different tasks. Over the years different models and theories have been presented to explain what triggers a task switch and how task difficulty level affects the switch.

Human action is directed by goals set in our minds. These goals result in tasks which aim to reach a predetermined goal. Salvucci and Taatgen [65] claim that two different conditions are responsible for triggering possible task switches. The first is an *external interruption* that requires immediate attention resulting in a task switch. Now the task hierarchy is re-organized and the interrupted one is resumed later. The second possibility is a *voluntary decision* to stop the ongoing task or take a break due to an obstacle that prevents the task completion. A new task is picked up to be continued with.

To apply these findings in a second screen scenario, attention drawing events targeted at any modality can be considered external interruptions. These could be the visual or auditive cue of an arriving tweet or the change of voice in the TV stimulus. Similarly, voluntary task switches can be triggered by boring TV or tablet content.

Altmann and Trafton [3] have introduced a theory to explain voluntary task performance and multitasking. It is called the *memory-for-goals*-theory and it defines a *goal* as "a mental representation of an intention to accomplish a task, achieve some specific state of the world, or take some mental or physical action". In order for a task switch to take place, the activation level of the goal at hand must be suppressed by the activation level of the rival goal in mind. This sets hierarchy for goals and task activation. The most recently activated goal determines the human behaviour.

Within the context of this thesis, the interest is in the self-interrupting conditions. Adler and Benbunan-Fich [1] have stated that there is not that much literature on discretionary task switching compared with the many studies on the interruptive counterpart. Though, Payne et al. [54] have found a tendency to voluntary task switch when the task at hand is no longer rewarding. In addition, Madjar and Shalley [39] report that people increased their creativity when they had an opportunity to switch tasks with specific goals. The switch allowed a break while continuing on the other task. With the increase in arousal they could focus better at the task at hand.

Adler and Benbunan-Fich [1] find voluntary task switching an interesting dilemma as to performance. On one hand, with a reasonable amount of task switches, people get additional stimulation and end up working

harder which leads to improved performance and focus. On the other hand, intensive task switching may make the environment distractive, which leads to decreased performance and excessive arousal. The scope of this thesis focuses more on free time applications and gratification. Performance issues are not that relevant.

It is worth while to notice that the switching times are longer when switching between high complexity tasks compared with switching between low complexity tasks as Rubinstein et al. [61] point out. As to the second screen viewing, the tasks can be considered as having quite a low complexity and task switch is thus quite inexpensive.

In previous research, Brasel and Gips [8] found that participants voluntarily switched attention between a computer screen and television on average 4 times a minute, in an experimental setup where participants were instructed to do whatever they pleased. They also found that gazes on computer screen lasted much longer than gazes directed on a television screen. In addition, younger people have a tendency to switch more frequently compared with older participants. However, other individual features such as multitasking preference and polychronicity (see Section 3.3.4) had little effect on how people behaved in the test situation.

2.1.4 Reasons for Multitasking

People have several reasons to do multiple tasks at the same time. Reasons for multitasking often relate to individual gratification, but also external factors, such as changes in media landscape, social norms and technologies encourage people to multitask [82]. These factors apply to media context as well. Here the viewing itself is most often related to seeking personal gratification and task switches are often influenced by external factors such as the content itself or social factors. As Montgomery [44] states, the media landscape is changing. Media is increasingly more important and integral part of the culture. The emergence of modern smart phones and social media is linking different media together and the changes "have spurred the proliferation of Web sites and other forms of new-media content".

The uses and gratification (U&G) model is a framework for the internal reasoning behind multitasking and is often used to describe media multitasking. The original paper by Katz et al. [29] lists cognitive gain and affective, social, personal and relaxing reasoning as drivers towards multitasking. These give plenty of reason for media consumption and media multitasking. For a more detailed view on U&G, see Section 2.4.4.

Viitanen et al. [82] summarize several studies that all agree with individual reasons for the U&G framework. In addition to the above it has been

reported that a non-media task can be made more multitasking friendly by adding media multitasking elements to it. Computer multitasking has also been explained with habitual use and convenience. The use patterns of human-computer interaction are explained in more detail in the work of Katz et al. [29]. The U&G theory has later been frequently applied to media research as seen in the work of Wang and Tchernev [83] and Ruggiero [62].

As explained by both Bluedorn and Jaussi [6] and Rideout et al. [60], many people also experience a gratifying feeling of increased efficiency when squeezing more media into to the same amount of time. Moreover, Ballagas [4] points out that the simultaneous use of both traditional and social media can also be seen as an outcome of people trying to accomplish traditional face-to-face communication over a distance.

Cognitive reasons have been found to be the foremost driving factor in multitasking. Wang and Tchernev [83] state that emotional needs may be indirectly satisfied even though this is not a goal while multitasking. This may even explain the tendency towards habitual multitasking.

In addition to the internal drivers, external factors also play a big part in media multitasking behaviour. McDonald et al. [42] have done a historical review of multitasking behaviour, explaining that as long as electronic media have been around, things have been used to escape boredom during household chores. First it was radio, followed by television. The mobile era has taken media multitasking to a whole new level. D'heer et al. [20] point out that tablets and smart phones are an integral part of many everyday activities. Followed by the statement made by Carrier et al. [11] this confirms that the growing trend is still continuing, the constant raise of electronic media is driving people to multitask.

2.1.5 Media Multitasking

Media multitasking is defined in literature [49, 79] as the act of finding and consuming multiple media simultaneously. In this thesis the same definition is followed. Further, in this thesis media defined as any channel through which content can be absorbed, i.e. news, entertainment, education, data or promotional messages.

Rideout et al. [60] have evidence of this phenomenon in practice. Studies have shown that media consumption have been increasing constantly over the 20th century. Now American youths consume media services 10 hours 45 minutes a day. Surprisingly the actual time spent in these media activities is just 7.5 hours a day. This means now more media is being consumed over a shorter period of time. In addition, the study reminds that the well known negative effects of multitasking on human performance

can also be seen in media multitasking.

Another viewpoint is provided by Ballagas [4]. He states that traditionally media multitasking has been thought of as very similar to any kind of multitasking, which means that every media is a separate task. This is, however, not always the case. Adding a new media to pre-existing ones cannot always be considered a new additional task from which to switch back to the pre-existing ones and back again. As stated by Ballagas [4], media multitasking is more about directing attention. That may even be an advantage in task solving and collaboration. This is why content creators and designers utilize media multitasking when introducing new designs, which leverage the advantages of the human multichannel media capabilities.

On the other hand, Novak et al. [49] consider that characteristics in media multitasking are the same as for any other multitasking situations with the exception that one media is usually considered as a primary media. The motivation to consume and process the secondary media is often to support the primary media. The possible distractive properties between the two media are less clear compared with other types of multitasking.

Another characteristic of media multitasking is the continuous simultaneous exposure to several information streams, which results in divided attention at the encoding stage. Therefore media multitasking is likely to decrease memory performance as explained in Section 2.1.1.

Reinsch et al. [58] introduce the term *multicommunicating* to describe multiple overlapping conversations. This phenomenon is increasingly more common in the hectic technology enriched 21st century media and office environments. It is also an indirect implication of ever increasing media presence Cameron and Webster [10], Reinsch et al. [58].

Reinsch et al. [58] claims that multicommunicating is an unintended use of new technology use (chat, email, etc.), that allow a flexible tempo in the communicating scheme. Social norms like productivity and efficiency, which encourage speed and interpersonal accessibility have made us more tolerant to divided attention and to a certain extent even delayed responses. Within some types of tasks and communicative situations this is even desirable to allow counter-parts to think about how to respond to the situations. Research also concludes that multi-communicating requires a special skill set. In addition to know how to operate the technical device, the user needs to feel the sensitivity of certain tasks and situations.

Multicommunicating skills can be considered essential in conversations taking place in situations where conversations span over multiple media or while consuming multiple media streams, such as a second screen TV viewing situation Cameron and Webster [10], Reinsch et al. [58].

2.2 Second Screen

The term *second screen* is used to refer to a computing device that is able to enrich the viewing experience that happens via a *primary screen*, most often a TV. Technological development has introduced a new generation of smart devices that allow broadcasting companies to take the viewing experience even further by providing applications to enrich specific programs. Sharing user-generated media has been an increasing trend and of human interest. Services such as YouTube, MySpace, Flickr are good examples. Following this trend and human social needs [89], social shearing of broadcast contents have increased. This has given TV viewing a new twist by taking off-screen activities to a whole new dimension.

Cesar et al. [13] have found four major second screen activities: *control, enrich, share* and *transfer*. Control refers to the use of the second screen as a control device of the media stream. Enriching means bringing in more content related media, such as commentary audio or additional information articles. Sharing denotes the social aspects and finally transfer refers to the plausible transition of the second screen turning in to the primary device. This thesis focuses on sharing and enriching, more specifically on the social interactions created by a second screen device.

2.2.1 Second Screen in a Media Environment

When studying literature it can be noticed that the concept of the second screen is not just a 21th century phenomenon. Cesar et al. [13] have found evidence of second screen TV viewing concepts dating back to 1996. In their review Cesar et al. [13] list several use cases for the second screen. The following use cases have been suggested: learning, content selection, providing content related extra information, electronic program guides (EPG), volume control and navigation, advertisement and commerce, and user participation (voting, chatting, affecting the narrative). Controlling the main screen with a second screen like device is prototyped by Choi et al. [16], who also presents encouraging results of a user study based on the prototype. Hess et al. [26] also recognise the interconnectedness of the devices. They state that a second screen gives easy access to enriching content related to the TV. Second screens are also capable of providing a personalized viewing experience.

The personalized media experience is often underlined. As a mobile device, the second screen also presents a possibility to enable viewing in situations previously impossible. Cesar et al. [13] also report that users

prefer to view different media on different screens, not to overload a single display while media multitasking. What has previously been done in the field of interactive TV, the solutions have only been limited to the use of on-screen display solutions.

Cesar et al. [13] have visualized the second screen environment as a media sphere. This sphere can consist of several devices: television, set-top-box, portable media player, mobile phone, a tablet or any other interactive device found in the home environment. Note the lack of hierarchy, this model does not place a device above another. One main concept is that these media spheres have communicative capabilities, enriching each other. To conclude, any handheld device with rendering and interaction capabilities can act as a screen.

2.2.2 Introduction of the Tablet — a New Environment

Müller et al. [46] found that after the introduction of tablets, tasks performed on mobile phones and laptops have been migrated and replaced by tablets in some use cases. Tablets are often used only at home while mobile phones are always carried around. Tablets are preferred over laptops because they are lighter and do not run hot. This is why tablets are replacing laptops as a second screen device and bringing computer use to environments previously not possible such as beds.

The same study [46] on tablet use reports that tablets, if used as primary screens, have a heavy tendency to be used for media viewing in activities such as reading books, news, watching videos and playing games. The three most frequent actions performed on a tablet are checking emails, watching videos and social networking. This is done while waiting but also while performing some other duties such as cooking. All in all, only a fraction of the reported tablet use is focused on work-related tasks. Sonera Oyj [72] carried out an interview study about Finnish media use and found that mobile devices are frequently used for TV viewing in situations where a traditional TV is not available, e.g. during holidays, when the TV is occupied, when a program is discussed with a friend, and in public transport.

The tablet also changes existing practices such us personalizing the viewing of broadcast TV, as explained by Hess et al. [26]. Müller et al. [46] mapped situations where the tablet was used for secondary activities. The most reported use case was to support watching TV. Participants of the survey reported that "their tablets can enhance the TV experience by extending that activity through, for example, looking up related information about the program that they were watching". This justifies the selection of a tablet device to be used in the experiment.

2.3 Increasing Media Consumption

The emergence of new smart devices and the availability of fast mobile internet connections has led to changes in media consumption. Traditional media has moved online and the use of social media has picked up its pace. This combined with increased use of mobile devices results in media being ever more present.

Studies (Rideout et al. [60]) and statistics (Statistics of Finland [74]) show that computer use has increased rapidly over the past few years. In Finland [50, 74] the growth has nearly saturated as yearly growth has lately been only marginal. The same sources show that changes in society have also led to more free time compared with what was two decades ago. This free time has provoked increased media consumption. Though challenged by increasing computer use, Finns still watch TV more than two hours a day. What is notable is that the role of a traditional TV set is challenged by computers, as a result of which viewing takes place increasingly via computers.

Statistics [51] show that the daily internet use in Finland is very active, but the growth is minimal as most people already have access to it. Some growth can still be seen in homes where new devices, such as tables and smart phones, and services, such as social media and transition to online newspapers, provoke internet use. Further, the role of the traditional desktop computer has been taken over by laptops.

2.3.1 Social TV

As with many other media, also television, has adapted itself to the 21st century changes. Rideout et al. [60] report increased sociability. This can be seen in the increased online content provided by broadcast companies and in the emergence of new types of TV shows (such as YLE Uutiset suora linja[1]) where viewers have a very active role. The program solely builds upon tweets sent around a timely topic. Viewers are also frequently endorsed to contribute. One way of doing this is to let viewers vote for today's good news to receive a few seconds of fame.

In addition, in the United States TV series have lately been provided with a hashtag to be used on Twitter in order for online societies to follow. These discussions, as reported by Twitter [80], often help viewers to discover new content and deepen further the engagement of viewers already familiar with a given TV program. In Finland, the national broadcasting company YLE even provides rules [88] for online etiquette.

[1] see Appendix A and `http://yle.fi/osallistu/suoralinja/`

As The Nielsen Company [75] states, the amount of online presence in a show strongly correlates with the number of ratings it gets. This correlation is most likely to be a result of the rise of second screen viewing. The effect is profitable for both parties: this can be seen throughout the season in the consistent views of episodes and consistent amount of tweets as well. In Finland the critical user mass is not yet to be reached regarding Twitter presence. Some arguable exceptions still exist like *Suora linja* which seem to have quality Tweets present online.

2.3.2 Interactive TV

Interactive television (iTV) refers to systems that make additional information available via a TV screen. The information may include, e.g., an electronic program guide (EPG). An interview study by Cruickshank et al. [18] brings forth how people perceive iTV. The study lists phrases like 'interacting with friends', 'more options', 'select and control', 'saving time', 'more information', 'personal service', 'giving you choices' mentioned by the test persons when asked to describe the situation. The early implementations featured in the study were designed in an on-screen fashion, i.e. the interactive menus were presented as an overlay on the primary screen. This blocked partly or fully the primary media. The usability of these systems was often frustrating: navigating through the content was often slow and clumsy, the remote control devices were not designed for proper interaction. Often the same information was available elsewhere e.g. in prints more quickly. These are considered as reasons why many iTV implementations failed, as summarized by Cruickshank et al. [18]. The tablet has provided a solution for increasing the usability, as explained in Section 2.2.2.

As explained by Simon et al. [71], it is possible to remove the conflict between the core functions of TV systems and the supporting auxiliary content, by using a second screen to display the iTV content. Thus it is possible to display additional content without disturbing the TV by occupying screen space on the primary screen. This second screen approach also provides a possibility to implement richer interaction mechanisms compared with classical implementations relying on a TV remote as the control device. Current second screen devices (tablets, mobiles) provide a much higher physical resolution and a unlimited portability. In addition, the second screen provides a more intuitive and pleasant way for the viewer to interact in comparison to the traditional infrared remote.

2.4 Social Presence

Media research and social processes provide us with theories to explain the key elements of social media. The *social presence theory* (Short et al. [70]) describes the media element of social media, stating that different media differ in "social presence". Here, the level of social presence is defined as the degree of visual, acoustic, and physical contact that can be achieved between the two communication partners.

According to [70], social presence can be considered as the product of intimacy and immediacy of the medium. Thus a high social presence is achieved in situations where intimacy is high and immediacy is low (e.g. face to face conversation). A lower social presence can be expected in highly mediated (low intimacy) situations such as a phone conversation or more asynchronous (high immediacy) situations as an email conversation. A higher level of social presence results in higher influence between communication partners and their behavior.

Another way to measure and classify media is to use the *media richness theory* as contended by Daft and Lengel [19]. This theory states that the goal of any communication is to resolve ambiguity and to clear uncertainty, thus differing media by its richness, of information it is capable of transmitting.

Goffman [25] has explained the concept of self-presentation where people have a tendency and a desire of controlling the impression other people form of themselves in face to face interaction. Electronic communication has brought along a range of new communication methods and developed a new communications etiquette, yet the same desire of control remains, as stated by Miller [43]. The controlled self-presentation is done to gain reward, e.g. to form a positive impression among friends, with the intention of forming an online image consistent with one's personal identity. This self-image is created through controlled self-disclosure as explained by Schau and Gilly [67]. Things that build up this image are thoughts, feelings, likes and dislikes people share while communicating. These are factors that can be considered as driving forces behind decision making in what gets shared on social media.

2.4.1 Social Media

As explained by Kaplan and Haenlein [28], the concept of social media is far from new. From its creation internet has seen many services allowing user created content to be made, e.g. Usenet, different bulletin boards and blogs leading up to social networking sites such as MySpace and Facebook

with others. The latter coined the term "Social media" and contributed to the prominence these services have today.

Mangold and Faulds [40] state that the last few years have been a massive expansion for social networking sites. They have penetrated to almost all parts of our everyday life. Now social media is also an integral part of companies' promotional mix. This has been adopted also by broadcasting companies by adding new social activities to support the viewing experience. The industry seems to have chosen Twitter as the social media of choice probably because of its simplistic nature of short and open messages with the possibility of tagging topics, which is ideal to go with TV viewing. For some time now, TV viewers have been provided with content related topics to promote online discussions and to gather interested people to join the discussion off-screen. Some channels have even made new program concepts almost fully dependent on what has been discussed online such as YLE Suora linja.

Marwick et al. [41] explain that people have a sense of audience even in online conversations and of presence dictating how they present themselves in these situations. Job interviews and night bars have all their norms and expectations. Also social media has its own imagined audience which people adapt our behaviour to.

Further, Thelwall et al. [78] state that Twitter may be a way to satisfy unrelated goals such as to create humour, show analytical skills or declare moral perspectives during TV viewing. As said in Section 2.4.4, the uses and gratification (U&G) suggest that people do not passively consume media but actively select and exploit it for their own goals.

2.4.2 Social Media Habits in Finland

According to Statistics Finland [50], four out of ten Finns contribute to some sort of content creation by writing on the internet. As online bulletin boards, blogs and commenting news sites draw most of the online activity, most of online content creation takes place in some sort of social media service. In the spring of 2012, half of the population in the age group of 16-74-year olds had registered on some social media site. Nearly all of those registered also follow those services. The most popular service is Facebook.

The use of these services is very age-dependent. The younger population is more frequently registered users and they also use the media more frequently. Out of the 16-24-year olds 86 per cent had been active in social media. The same number in the group of 55-64-year olds was only 22 per cent. What comes to social media, the behaviour is not notably different

between the sexes.

The same age-dependent trend is seen in the United States and the UK in a study by Lenhart et al. [37], where Facebook is used more actively by young users. However, Twitter is the choice of elder people. The same trend might well apply to Finland as well, as explained by the lack of Twitter activity. This is further explained by the small national population, resulting in a small pool of Twitter messages per program.

2.4.3 Twitter

Twitter is a well-established micro-blogging site as explained by [15]. The messages posted (*tweets*) in this service are short, limited to 140 characters. The service is considered as a social networking site as each member has its own profile and members can be "friends" with other people by following them. Regardless of that, all the messages posted on the site can be seen by everyone.

Television and Twitter are frequently combined [14, 15, 75, 78, 87]. Both domestic and international broadcasting companies consider Twitter as the de facto social media platform for TV, most likely due to the nature of compact and public messages. Twitter can be considered as an industry standard and a good reason for choosing it to represent social media in our experiment as well.

In order to understand the Twitter environment, it is essential to know some terminology associated with the service. Firstly, *hashtag*, identified by "`#`", gathers all tweets associated with a specific word. E.g., "`#thesis`" gathers all thesis related tweets. This meta information is used to mark topics, helping people to navigate and follow the topics. This also emphasizes the typical features of wide communication.

Secondly, *user mentions*, are represented by `@`-prefix. With user mentions, members can address other registered users of the service.

Finally, *retweeting*, as explained by [78], is a method of disseminating a tweet by reposting it. The aim of retweeting is to spread a tweet the user found valuable or worth highlighting for some reason. If a tweet goes viral, the retweeting of this tweet can be very rapid.

2.4.4 Modelling Needs and Fulfillment

The uses and gratification (U&G) model is a traditional way of modelling human media use. U&G tries to answer *why* people do what they do with media instead of what media does to people ('media effects'). Media research has been using U&G in a broad manner trying in general to answer

the question *what* people do with media, allowing a range of interpretation and responses.

As explained by many [29, 35, 62], U&G originate from social sciences, and it presents the use of media to fulfill personal social and psychological needs. Gratification can be obtained from a specific content or medium (TV programs), media (social media, newspapers etc.) and social content (watching TV with friends).

U&G argues that individual needs specify how people use a specific media they have chosen. Bryant and Zillmann [9] have noted that mood influences the choice of media. Stressed people are likely to choose a relaxing content and bored people are likely to choose an exciting media. The same media may gratify different needs in different people.

A typical U&G study is likely to use self reporting data to explain *why* the observed behavioural patterns appear. This is why the U&G methodology has also raised some criticism. Viewers might not now why they did the choices they did. The answers might also not be that clear and influenced by explanations heard from others. Media is often forced upon people and in this way not conscious and active choices that they are exposed to [79]. This research indicates that media use can be used as a result of addiction.

Chen [15] describes U&G as a model for media use. A need is defined as disequilibrium of internal and external occurrences that strive for equilibrium. That is, if people have a need, they seek to gratify it. U&G means that multiple media compete for human attention and that people select the medium that best meets their needs, which might be a need for information, a desire for emotional connection.

In addition, Chen [15] lists several studies where U&G has been successfully used to explain behaviour on the web, social media and media use. This is supported by the fact that U&G strives for a model on what people do with media, not what media viewing does to people. If people are given a possibility to pick any media and to stay with it, U&G states that it must satisfy their needs.

With the above and the interruptive customs of second screen explained in Section 2.1.2, it can be argued that people do the things they prefer and which give them the greatest satisfaction possible.

2.4.5 Current Implementations

Some commercial second screen applications have already emerged and the user base seem to slowly expand. PC Magazine [84] listed the 5 greatest second screen apps in late 2013. Moreover, Evolumedia group [23] has also

done a survey on the present companion applications. According these sources, the three most widely used second screen applications after Facebook, Twitter and Shazam[2] were Yahoo IntoNow[3], GetGlue[4] and ZeeBox[5].

The two first mentioned had at that time 3 million users. ZeeBox, with a million users in late 2012, was estimated to reach 6 million users after its American introduction. America really seems to be the home of second screen viewing, as most of the companion applications mentioned here operated in the United States with the exception of ZeeBox originating from the United Kingdom.

Features included in these companion applications range from audio based program recognition to rewarding the users with virtual prizes, program related social content feeds, and lately even meta data listings. Most of the programs are available on the majority of the mobile and tablet platforms which include Android and iPhone. ZeeBox feature support Amazon's Kindle e-reader and a web interface accessible via a browser.

These are generic companion applications working together with many TV programs. In addition to these, there are program specific second screen applications. These dedicated applications feature the most customized and targeted content. It is also the most expensive way of reaching multitasking TV viewers. On the other hand, the dedicated applications offer the best possibilities for monetization via advertisers and sponsors [23, 24].

The companion application used in the experiment was designed to mimic the features of a more common companion application. Program specific content was provided through medias already known by the participants. These details are presented in detail in Section 3.1.3.

[2] A music-recognition application
[3] `http://www.intonow.com/`
[4] `http://getglue.com/`
[5] `http://zeebox.com/`

Chapter 3

Experiment

This chapter describes the design of the experiment. The experiment was designed in order to answer the research questions presented. As controlling the second screen content was essential, this requirement was taken as the basis for the design. The display of the second screen content was based on four second screen content conditions. The stimuli used consisted of both TV and second screen content. Eight video clips of nationally broadcasted television was used as the main screen content. And as the second screen stimuli, tweets and articles of additional information related to the program were used. This made a total of four different second screen content variations as listed below. If present, the second screen content was displayed throughout the whole trial.

I TV content + 2nd screen: no content

II TV content + 2nd screen: additional information

III TV content + 2nd screen: social media

IV TV content + 2nd screen: as in II and III combined

Out of the conditions listed above four *condition sets* were gathered. These condition sets consisted of a predetermined combination of the four possible second screen conditions. This way one participant was exposed to all of the possible second screen content conditions during the experiment. Each participant was shown the stimuli in a randomized order according to the stimuli combination determined by the condition set. With this design all stimuli combinations were shown after every fourth participant. Ravaja [57] has given advice on how to conduct psychophysical media studies and these guidelines were pursued.

The experiments were performed in a room dedicated to visual testing at the Department of Media Technology. This is a controlled environment,

that has previously been used to perform similar experiments within the department.

In between the stimuli each participant did self-reporting (see section 3.3 and appendix C for details) on how they experienced the situation. Additionally, in order to capture both objective emotional information (arousal and valence) and the distribution of attention between the two devices each participant was monitored for psychophysiological signals in addition to wearing eye-tracking glasses.

The additional information mentioned was used as stimuli for an other thesis done in parallel with this work. This other thesis [31] studied how accompanying fact-based articles on the second screen affects the viewing experience. The additional information provided consisted of Wikipedia and YLE news articles.

3.1 Experiment Design and Environment

The aim of this section is to lay out in detail the experimental set-up. The following subsections will present the devices, environmental conditions and software used. Figure 3.1 shows a participant seated during the experiment.

3.1.1 Hardware

In the experiment a Samsung Galaxy Tab 2 Android 4.0 -tablet Samsung Electronics Co., Ltd. [66] was used as the second screen device. This model had a screen size of 10.1 inches with a resolution of 1280×800, resulting in a pixel density of 149 ppi. When the device showed total black, the light emitted from the display was 0.20 lux, while the full white the amount of light emitted was 130 lux.

In order to alleviate the later analysis of the Eye Tracking Glasses (ETG) data, the tablet was fastened in an adjustable table mount in front of the subject. Upon introduction, the participants were asked to adjust the tablet screen to enable pleasant viewing, but at a minimum distance of 40 cm as dictated by the limitations of the eye-tracking glasses [69].

A Sony 40" FullHD (1920×1080) LCD-TV (Sony KDL-40HX800, 240Hz) was used as the main screen. Illumination readings of 0.0 lux was measured when displaying black and 160 lux showing full white. The TV was stationed at a distance of approximately 200 cm depending on how the participant chose to sit on the chair. The rooms light was adujsted to accommondate pleasent viewing of both screens. This complied with the ITU rec-

ommendations International Telecommunication Union - Radiocommunication Sector [27] for visual testing.

Psychophysiological data was collected using Varioport, a portable biosignal recorder by Becker Meditec, Karlsruhe, Germany. This device is capable of recording EMG (Electromyography), EDA (Electrodermal Activity) and HR (Heart Rate) signals at a 16-bit resolution and a sampling rate of 2 kHz. Gaze tracking was done with SMI[1] Eye Tracking Glasses. These glasses carry tree cameras: two capturing the pupil movements at a sampling rate of 30 Hz, the third capturing the scene at 24 frames per second.

Figure 3.1: A participant seated during experiment. All sensors and both the tablet and TV can be seen in the picture. The tablet was mounted in front of the participant for easier gaze tracking.

3.1.2 Environment

The room was lit with fluorescent ceiling lights varying between 73 lux (both displays displaying black) to 125 lux (both displays showing white). These measures where taken from where the participants were seated, facing in the direction of the screens. The walls of the room were covered

[1] SensoMotoric Instruments Ltd.

with mid gray curtains to ensure a neutral viewing environment. The room was located by a peaceful corridor with all windows blocked with curtains. Signs indicating a ongoing experiment were posted outside instructing bypassers to keep quiet. To ensure a completely distraction-free environment the participants wore headphones to block occasional loud noises coming from other parts of the building.

3.1.3 Software

As stated previously, the test environment consisted of several devices that all needed to be synchronized for precision timed stimuli display and data recording. A desktop computer controlling the experiment procedure served as the main hub for all the separate devices. In addition to controlling the display of TV and tablet content, it also recorded participant interactions with the tablet and the data sent wirelessly by the Varioport measuring device.

Furthermore, a laptop specifically set up for recording the data from the eye-tracking glasses was also used to access the experiment manager. The post experiment questionnaire (appendix D) was filled in on this laptop. The well planned and executed implementation of the experiment manager allowed easy and consistent trials. Only the eye-tracking software needed to be manually operated because of the lack of a sufficient Application Programming Interface (API) access.

The software allowed us to meet the requirements set by Ravaja [57] regarding research involving psychophysical measurements. Strict requirements apply to the synchronisation of stimuli presentation and the recording of the psychophysical data. To allow analysis between many different measures there must be a way to precisely link psychophysical events with each other and the events in the stimuli.

All of the above was hidden from the participant, who only saw the tablet interface depicted in Figure 3.2. Efforts were made to make the second screen application as user friendly as possible in accordance with familiar use patterns seen in many existing touch screen applications and in accordance with Nielsen's [47] usability guidelines.

New content appeared on top of both the tweet and the article feeds as older items were shifted out of sight at the bottom of the screen. Each article could be opened by pressing the items now appearing in full length in the middle of the display area. Tweets were rendered in the same manner as they appear on Twitter's website[2]. Tweets could also be re-tweeted in the same fashion as in the real service.

[2] `http://www.twitter.com/`

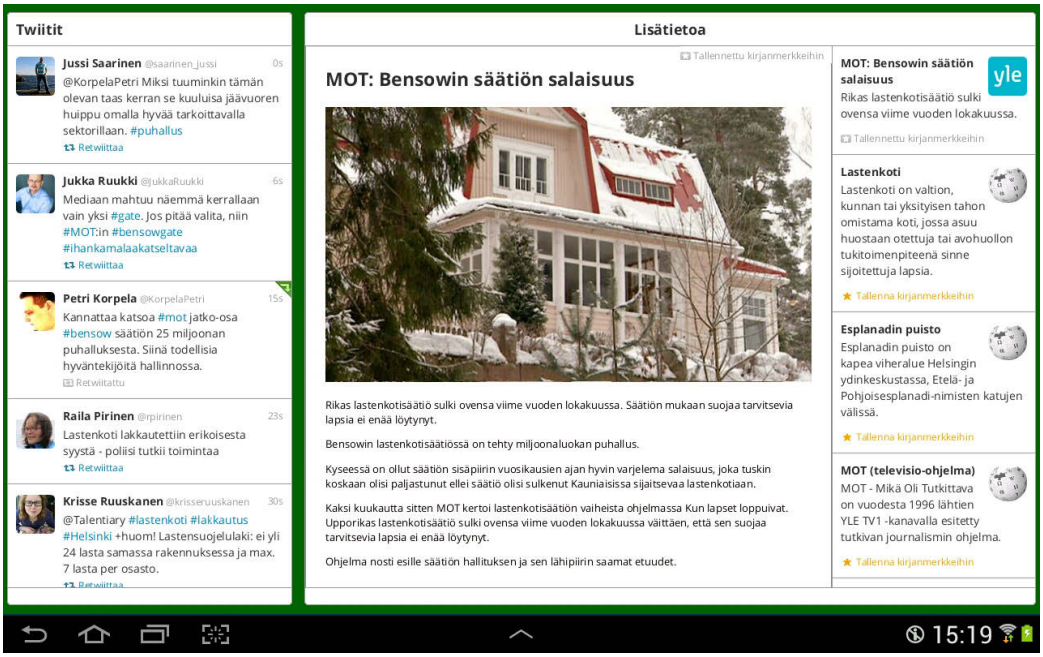


Figure 3.2: Screenshot of the user interface as shown on the tablet device when both tweets and articles are displayed. Tweets can bee seen listed on the left. Additional information is on the right on topic level. Once opened individual articles can be freely read in the middle of the screen.

3.2 Stimuli

To answer the research questions presented in the introductory Chapter 1 as comprehensively as possible a set of four program genres were chosen: magazine, documentary, reality and sports. These could further be grouped into two, fact and entertainment. Two programs were chosen to represent each genre. This summed up to eight different stimuli. In the stimuli selection, efforts were made to make the range of program pace and intensity as wide as possible. This way the effects of the second screen application could be monitored in as many situations as possible. Clip length was chosen to be around 5 minutes, long enough to allow sufficient engagement into the stimuli and to be sure to reliably capture psychophysiological measures. The five minute time frame is considered of sufficient length by previous experiments [33].

Stimuli selection was to some degree dictated by the low amount of TV related tweeting in Finland and the fact that Twitter search API only provided a shallow access to the tweeting history. This led to TV content

being selected based on the number of twitter messages available once the predetermined tweet gathering period had finished.

As this fact set the order of procedures, a preliminary list of programs was put together with the help of the TV schedule and the program guide, followed by figuring out appropriate keywords to use for finding program specific tweets. With the keywords list ready, collection of tweets could be started with scheduled polling script.

There are some programs that provide a hashtag for user interaction on Twitter. This enables easy on-line discussions to form around a program. This also enables the broadcasting company to build programs tightly linked with viewer tweets. Examples of these are ”Pakko Tanssia”, a dancing contest, and ”Suoralinja”, an interactive news program based on strong viewer interaction. Respectively, there are more traditional programs with almost zero activity on Twitter (documentaries). Yet even programs with strong Twitter promotion got very few tweets directly linked with the program. It is thus clear that Finnish tweeting habits are far behind in comparison with tweeting habits in countries with a larger population. It remained unexamined whether the reason for this is the population or the tweeting habits.

Table 3.1: List of TV stimuli used in the experiment. A more detailed listing can be seen in Appendix A.

Label	TV program name	Genre	Group
`bensow`	MOT: Bensowin säätiön salaisuus	Magazine	Fact
`suoralinja`	YLE Suora linja	Magazine	Fact
`hanhet`	Avara luonto: Stadin valkoposket	Documentary	Fact
`rakkaus`	Totuus rakkaudesta	Documentary	Fact
`latela`	Latela	Reality	Entertainment
`pakkotanssia`	Pakko tanssia	Reality	Entertainment
`robbe`	Boxing: Robert Helenius vs. Michael Sprott	Sports	Entertainment
`salibandy`	Salibandyn EFT: Finland vs. Sweden	Sports	Entertainment

Following a tweet collection period of a few weeks, it was agreed which

programs to show, mainly based on the tweeting activity. Table 3.1 lists the selected TV programs. A more detailed listing can be found in the Appendix A. Label presented in the above table are used throughout the text to compactly refer to each program.

Partnering in the NextMedia research project, the national broadcasting company Yleisradio (YLE) provided the stimulus used in highest possible broadcast quality. As some programs were delivered in high definition (HD) and some in the lower resolution standard definition (SD) all video clips used as stimulus were rendered into SD.

In order to understand what effect the stimuli have on the participants, reference affective values were needed. Thus both tweets and television programs where validated as explained in the following sections.

3.2.1 Video Stimuli Ground Truth Collection

Baseline affective values (the ground truth) of the stimuli were needed in order to compare the different experiment conditions. For this purpose a tool was made allowing simultaneous capture of valence and arousal values over the temporal dimension. In this thesis, the action of collecting the ground truth data is referred to as *validation*.

This tool combined the two SAM (see Section 3.3.2) scales into one two dimensional area, as shown in Figure 3.3. The method allowed free movement across both scales. In order to lower the cognitive load familiar smiley faces were selected for reference in each quarter. The persons doing the validating were instructed to refer to original SAM guidelines for emotional definitions and the use of the scales.

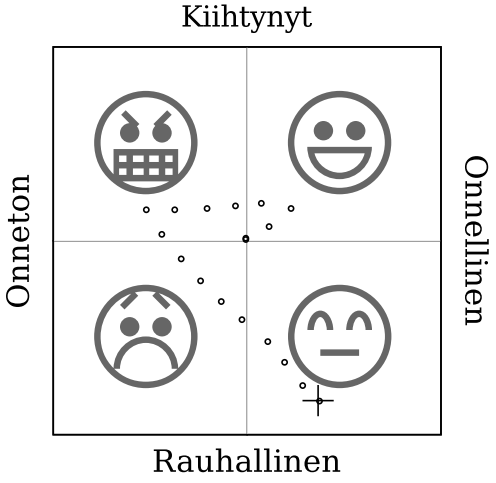


Figure 3.3: The two-dimensional validator interface. Arousal was measured on a scale from aroused (kiihtynyt) to calm (rauhallinen) and valence from happy (onnellinen) to unhappy (onneton). The cursor is depicted as a cross-mark, the dots are recorded values.

The tool was implemented as a website, and it was accessible online. This gave the validators a possibility to choose the validating sessions freely over the course of a few weeks. Validators were instructed to assess each video according to what they felt during viewing. Video playback started when the mouse button was held down in the validating area. This choice was made to make sure that the person validating was alert and focused on the task. The window playing the video was situated right above the validating area. Videos were played in a random order to each validating person.

Each video stimulus was validated with this tool by five co-workers from the research group. After visual inspection of the plotted data this was considered a large enough sample group for the purpose at hand to consider results trustworthy. For the outcome, see section 4.2.

3.2.2 Tweets

By the time of writing, Twitter.com search API only offered a shallow access to the tweet history. This practice forced to develop methods for capturing tweets over a period of time. A set of command line tools was written to enable automated capture and storage of tweets.

A job scheduler ran the tweet collector every hour to collect new tweets and stored them in a local database with all the available meta data associated with each individual tweet. This method allowed us to monitor the accumulation of tweets for each candidate program and to choose the final TV stimulus at a later stage.

Once a list of candidate TV programs had been chosen to use as stimuli, a list of hashtags associated with each program was prepared. Some of the TV content did not offer a hashtag to go with the program, this led us to choose hashtags we felt best represented the programs. This list of hashtags was fed in the tweet collecting tools mentioned above.

Xu and Croft [85] describe a query expansion method where a query is initiated with a set of keywords followed by a query expansion using an expanded set of keywords found in the results of the initial query. The final set of tweets was gathered with this method in cases where the first query round did not result in a sufficient amount of tweets. In all cases two query rounds resulted in a sufficient set of tweets.

A chose was made to unify the tweets by selecting only tweets written in Finnish to remove the possible complication of multiple languages. Also the tweets containing images were ruled out and any external URLs were removed from all of the selected tweets. Finally tweets that were the least related to the program were manually removed. Further manual pro-

cessing was needed when a similar tweets per minute ratio was tried to be reached for all the programs still paying attention to the natural genre specific differences. Similarly, the tweets least related to the TV content were removed.

3.2.3 Sentiment Analysis of the Tweets

Previous research (Laine-Hernandez et al. [33]) has found a significant effect of tweet sentiment on the viewers during TV viewing. Negative tweets can lower the viewing experience of positive content as well as positive tweets can improve the experience of negative content. This in mind, in order to interpret the data collected in the experimental phase, the tweets being part of the stimulus, a sentimental analysis was required.

As summarized by Thelwall et al. [77], literature knows several terms used to describe different types of emotional analysis. The term *sentiment* refers to splitting emotions into positive, negative and neutral. This should not be confused with *emotion* which refers to the affect of sentiment (e.g. happy, sad etc.). As explained by Thelwall et al. [78], a third term used in literature is *Opinion Mining*, which refers to finding sentimental opinions from unstructured text, for applications such as determining movie popularity through online comments. Opinion mining is usually just meant to detect sentiment (negative or positive), not the strength of the sentiment. Opinion mining also refers to more advanced applications such as to identify if statements are subjective or objective.

Initial tweet analysis was done with SentiStrength (Thelwall et al. [77]), a tool developed for short online messages. The algorithm, initially developed for English, had a Finnish translation made by Dr. Kakkonen [3].

SentiStrength was developed to classify MySpace[4] comments. These comments contain informal language such as slang and abbreviations. To exemplify creative online spelling the messages may include emoticons like `:-)` and stretching of words by misspelling them (e.g. ”haaaapppyyy”) giving these a higher valence. These characteristics are also frequently present in tweets. Thelwall et al. [77] found that SentiStrength was able to classify the messages with sufficient accuracy.

To score sentiment SentiStrength introduces a five-point scale for both positive and negative sentiment. The final sentiment score is derived from these two figures. The algorithm reports values for negative and positive

[3] Dr. Tuomo Kakkonen, Joensuu University, Finland `http://cs.joensuu.fi/~tkakkone/`

[4] A social networking service with a strong music emphasis

as follows: a negative sentiment is scored from -5 (extremely negative) to -1 (neutral) and positive sentiment from 1 (neutral) and 5 (extremely positive). The highest score in both negative and positive is reported as the final score in each category. Thelwall and Buckley [76] explain that in situations where a single number is needed to represent the whole sentiment the positive number can be added to the negative value resulting in a sentiments scale from -4 to 4.

Later inspection and quality control exposed the weak performance of the SentiStrength algorithm. This resulted in manual annotation. As there were over 400 unique tweets we had only the resources for one individual to do the annotation. In a similar fashion to SentiStrength each tweet was given a negative and positive sentiment on an absolute scale from 1 to 5.

3.3 Collection of Data

In oder to capture a comprehensive picture of the *viewing experience*, a wide range of meters were used. This set of both objective and subjective meters are listed in Table 3.2. The table also lays out the phase in which each of the individual metrics were utilized: prior to, during or after the experiment. All these are presented in further detail in the sections to follow.

Self-reporting is a viable tool while researching subjective media experiences. Many studies [33, 46, 63] have used self-reporting questionnaires for data collection. Also U&G methodology relies heavily on self reporting done by participants.

Likert-style reporting was used to capture the opinions during the test. In addition, measuring emotions was an important part of this study. This is why Self Assessment Manikin (SAM) was selected to be the self-reporting tool of emotion. A complete list of the questionnaires used can be found in the appendices. The adjectives used to describe the stimuli in the results chapter were derived from the questions presented in the questionnaires.

Table 3.2: List of techniques used to measure the participants' response.

Meter	Type	Phase
Background questions (Appendix B) *Gathered basic information and media consumption habits of the participants.*	Subjective	Before
Polychronicity index (Appendix B) *Used for measuring the individual multitasking preference.*	Subjective	Before
BIS & BAS index (Appendix B) *Used for measuring the participants' individual behavioural tendencies.*	Subjective	Before
SAM questions (Appendix C) *Measures arousal and valence after the each trial.*	Subjective	Post-trial
Post-trial questionnaire (Appendix C) *Qualitative questions regarding both tablet and TV content.*	Subjective	Post-trial
Number of Tablet Interactions *The interactions made measure level of involvement with the tablet.*	Objective	During
EMG-OO *EMG response of Orbicularis Oculi, measure of positive emotion.*	Objective	During
EMG-ZM *EMG response of Zygomaticus Major, measure of positive emotion.*	Objective	During
EMG-CS *EMG response of Corrugator Supercilii, measure of negative emotion.*	Objective	During
Heart Rate *Heart Rate was used as a measure of arousal.*	Objective	During
EDA *Electrodermal activity was used as the measure of arousal.*	Objective	During
Eye-tracking *Measures distribution of the participants' visual attention.*	Objective	During
Post-test questionnaire (Appendix D) *Captures an overall picture of participant sentiment: free comments about both companion application and experiment.*	Subjective	After

3.3.1 Likert-style Scales

Croasmun and Ostrom [17] describe the use of Likert-style reporting. Likert scales are broadly used as a cross disciplinary answering scheme. The same study shows that the use of a Likert-style scale is a very useful and relatively liable tool for self-reporting. In Likert-style reporting the participant is asked to respond to statements with a scale ranging between two extremities, usually ranging from "strongly disagree" to "strongly agree". Usually there are five categories to choose from, with a neutral choice in the middle, sometimes expanded to seven or nine in order to provide increased accuracy or freedom of choice. Some researchers argue that a seven point scale should be used for optimal reliability. More points than seven would not increase reliability. Even-numbered scales force the responder to take a certain position, even if no verbal opinions are presented. However, the number of points should always depend on the situation. Negatively worded questions and answers should be avoided as they increase cognitive load.

For internal consistency and reliability it is essential to calculate Cronbach's alpha. This refers to how well the measurements and instruments worked and it measures the liability of the measured results. It is not a statistical measure but a measure of reliability (consistency).

3.3.2 Self Assessment Manikin

Human emotion is a subjective thing that has early been measured via lengthy interviews till the development of the Self Assessment Manikin (SAM) system. SAM is the work of Bradley and Lang [7], and as explained by the authors, SAM is a "non-verbal pictorial assessment technique that directly measures the pleasure, arousal, and dominance associated with a person's affective reaction to a wide variety of stimuli."

Results gathered with SAM correlate strongly with emotional measurements gathered via verbal methods. Thus, SAM is ideal the capture of the emotion of a participant after the exposure to a stimulus. The non-verbal method allows the use of SAM even for non-English audiences, children etc. Research has proven SAM to be a reliable tool for measuring emotions for a wide range of stimuli.

Here, within the scope of the thesis, it is justified to only inspect two of the three major affective dimensions. Experienced *arousal* (exciting/not exciting) and *valence* (pleasantness/unpleasantness) is measured with SAM. As said, valence can be considered as the pleasure dimension, while arousal represents the wakefulness or excitement level.

The third dimension on the SAM scale is *dominance* (how well in charge of the situation you feel you are), which can be assumed does not play a big role in a home media viewing environment. It is not frequently seen in TV viewing studies, either.

Measuring arousal and valence is reasonable, as these are considered to account for the media experience. Stronger values equal a stronger effect and lesser absolute arousal and valence readings result in a more diluted experience.

3.3.3 Behavioural Inhibition and Activation

Carver and White [12] summarize that behaviour and affect are controlled by two underlying systems: the Behavioural Inhibition System (BIS) and the Behavioural Activation System (BAS). *Gray's theory of Brain Functions and Behaviour* states that BIS can be considered the aversive motivational system and that this is the mechanism that controls the experience of anxiety in response to anxiety-relevant cues. BIS is said to inhibit actions that might lead to painful or negative outcome. BIS is also linked to the experience of fear, sadness and frustration. Conversely, activation of BAS drives humans towards activity leading to goals of positive outcome and the experience of elation, hope, happiness and other positive feelings.

Thus, BIS/BAS index is an expression of human personality. People with high BAS should react behaviorally to situations with positive cues and experience positive affect. Vice versa, people with high BIS should experience a negative affect in a situation with cues of punishment and negative outcome more than people with a lower BIS index.

In order to measure BIS and BAS, Carver and White [12] developed a self-reporting questionnaire that can be used to determine the individual level BIS/BAS sensitivities. A translated version of these Likert-style questions used in the experimental phase are listed in Appendix B.

3.3.4 Polychronicity

Polychronicity is a term used to describe human preference of working on multiple things at the same time. A simple example of polychronic behaviour is to prefer to cook and watch television simultaneously as opposed to monochronicity, the preferred behaviour of doing only one single thing at the time.

Polychronicity, as explained by Poposki and Oswald [56], is a measure of individual preference for multitasking opposed to performing a single task at a time, whereas multitasking refers to the behavioural aspect of

polychronicity. Though, König and Waller [30] remind that literature still lacks a clear definition of polychronicity.

Poposki and Oswald [56] explain that a tendency to multitask is likely the result of positive multitasking experiences and the tendency to perceive multitasking as efficient and rewarding. Thus, polychronicity can be considered a useful indicator of multitasking-related constructs. Though critics (König and Waller [30]) point out that studies exploring this theory have got quite mixed results even though the idea is backed up by both theoretical and logical sense.

Poposki and Oswald [56] introduce a *Multitasking Preference Inventory* (MPI), a reliable tool for measuring polychronicity. MPI consists of a set of scored questions whereas aggregated results in a figure represent the tendency to multitask. MPI was part of the pre-trial questionnaire, see Appendix B for Finnish translations used by us.

3.4 Psychophysiological Measures

According to Bradley and Lang [7], emotional reactions can be measured at least with physiological reactivity, affective reports, and overt behavioral acts. They also maintain that an emotional state is a sum of experienced *arousal* (exciting/not exciting), *valence* (i.e. pleasant/unpleasant), and *dominance* (how well you feel you are in control of the situation).

Ravaja [57] introduced the idea of utilizing psychophysiological measurements within media research for measuring attention and emotion. Three most commonly used psychophysiological measures are heart rate (HR), facial electromyography (FEMG) and electrodermal activity (EDA). The same publication argues for the use of psychophysiological measurements in affective research. First, many of the psychological phenomena (attention, emotion and arousal) have an equivalent in the psychophysical components (a response triggered by the autonomic nervous system). In addition, psychophysiological measures may provide information that goes unnoticed by other methods like observation or self-reporting.

Second, psychophysiological measures can be considered more objective than self-reporting, as self-reporting may be influenced by social norms when researching sensitive topics such as political matters.

Third, psychophysical measures enable continuous measurements during stimulus exposure. This results in multiple data points which enable inspection of time series and a more elaborate analysis than a single data point self-reported value. In addition with psychophysiological measures it is possible to observe and collect data without a person's active cooperation.

With many eligible characteristics there are still some things to remember. Ravaja [57] brings forth that psychophysiological measurements in media research rely on theories and constructions developed around psychophysical findings that are based on traditional simple stimuli (e.g. tone beeps). These simple stimuli are way more simple than the multimedia stimuli used in media research mimicking real life situations.

Research done by Novak et al. [49] found that psychophysiological measures are not easy to link to a specific task during a multitasking session. The study also states that it is relatively easy to obtain information about the human physical state. However; it is far more challenging to process these measures into a value representing a subjective feeling, although the assumption is that information about the human mental state can be gathered via capture of psychophysiological measures. Furthermore, results indicate that psychophysiological measures are sensitive to mental workload but are not always agreeing with the subjective feeling.

3.4.1 Measures of the Human Nervous System

In order to understand the reasoning behind the psychophysiological measures preformed, it is necessary to briefly look at how the human nervous system works and what causes the measurable reactions.

According to Encyclopædia Britannica Online [22] and Noback et al. [48], the human nervous system is a system that is responsible for adjusting the internal human balance to best cope with the external environment. In other words, it carries stimuli received from sensory receptors to the brain and back to different parts of the human body. The nervous system can be divided into several subsystems. Here the main interests are in the *Sympathetic Nervous System* (SNS) and the *Parasympathetic Nervous Systems* (PNS) as they control the physiological responses measured during the experiment. Both these systems are part of the Autonomic Nervous System (ANS) which is involuntary. ANS controls many of the physiological functions such as heart rate, respiratory rate, glands (e.g. perspiration) and arousal.

As described by Noback et al. [48] the SNS stimulates activities in a stress situation. The SNS is responsible for an increase in heart rate, blood pressure and sweat gland stimulation in high arousal situations. The PNS is, in contrast, responsible for restoration and conservation of body resources. This includes decreased heart rate and increased digestion. PNS is activated in tasks demanding concentration.

3.4.2 Heart Rate

Referring to Ravaja [57], heart rate (HR) is measured as number of heartbeats per unit of time. HR is most often measured by electrical potential generated by the heart during a cardiac cycle. Thus the term electrocardiography (ECG). In the experiment, HR data was collected with the Varioport and a sensor placement depicted in Figure 3.4.

Changes in HR depict many of the changes within the human body. To list some: attention, effort, arousal, and emotion. All of these are frequently used in media research. An increase in HR is associated with arousal, preparation for action. In contrast, information intake and tasks requiring attention, such as consuming media, is known to lower the HR.

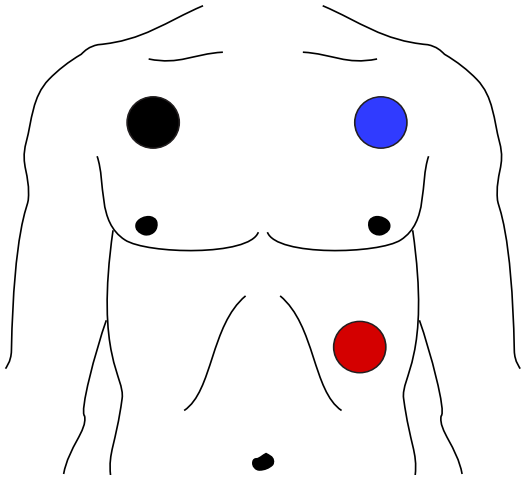


Figure 3.4: ECG electrode placement according to the Einthoven's triangle, a reliable way to measure the R-spike.

Allen [2] claims that HR is not a univocal indicator of emotional arousal, meaning that HR is not a very good indicator of emotional arousal as it is influenced by both the SNS and the PNS. This must be taken into account while interpreting HR readings. Nonetheless, HR can be used as a measure of attention but the interpretation must be supported by other data [5, 57].

Allen [2] states that the most commonly extracted indicator is the interbeat interval (IBI), which is the time in milliseconds between two sequential R-spikes. R-spikes are well identifiable as they are the most prominent feature of the heartbeat. The IBI series are frequently used as input of algorithms that compute metrics based on HR measures.

3.4.3 Electrodermal Activity

Electrodermal activity (EDA) and galvanic skin response (GSR) give a measure of the electrical conductance of human skin. This is related to the level

of perspiration, which in turn is controlled only by the SNS. Thus, changes in EDA are used in media research to measure arousal.

Ravaja [57] reports that arousal highly correlates with self-reported levels of emotional arousal. EDA can be used to measure several different processes: activation, attention, and task significance as experienced by a subject.

As explained by Ravaja [57], there are two types of features that can be extracted from the EDA activity: tonic and phasic. Tonic or gradual changes are measured and reported as the skin conductance level (SCL), while the skin conductance response (SCR) is used to measure the phasic, rapid changes. More specifically, SCRs are temporary increases in the measured value, soon followed by a return to a tonic level. SCR events are orienting responses to strong momentary stimuli. The tonic EDA (SCL) is used to measure long term changes in arousal.

Both Schmidt and Walach [68] and Ravaja [57] describe the tonic EDA or SCL as follows. Low SCL is associated with low arousal and vice versa. Low arousal is associated with pleasantness, joyfulness and relaxation. This can be achieved with peaceful stimuli and low rate of cuts in the edit. In contrast, high arousal is a result of increased irritation, anger, fear (depending on the dominance level) and decreased levels of pleasantness. This could be a result of arousing stimuli such as deliberately designed radio or TV messages with a high rate of edits or a situation in which the content is presented on a large sized screen.

As Novak et al. [49] lay out, the SCL is the basis for interpreting skin conductance measurements and a basis for interpreting SCR. Baseline values for these measurements can be obtained from EDA measurements conducted while no discrete environmental events are present. A mean value is calculated over a distinctive predetermined time. This is considered to be a zero value for the experiment period, as many of the instruments do not measure absolute values, but only change from the initial values. The SCR peaks are extracted from the SCL signal with algorithms developed specifically for this purpose.

As pointed out by Ravaja [57], the criticism against the use of the EDA is the fact that it is difficult to distinguish individual psychological processes from each other. This underlines the importance of stimuli selection and control of the experimental conditions. Another thing to point out is that stimuli in media studies are often relatively long and calm, and the EDA response decreases due to habituation rather quickly.

Schmidt and Walach [68] instruct that when placing the EDA sensors the non dominant hand is usually preferred, as movement has an influence on the measurements. Further the palm side of the hand is used where

there are two possible locations to place the sensors. These locations are shown in Figure 3.5. The electrodes can be positioned on medial phalanges or the thenar and hypothenar eminences. Constant good contact is crucial, and is why it is important to make sure that the electrodes are fastened securely. It was decided to position the electrodes on medial phalanges of the little and index fingers as depicted in Figure 3.5.

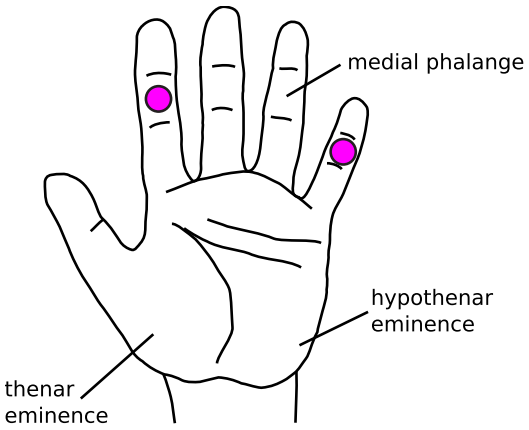


Figure 3.5: Possible locations for EDA electrode placement. Electrode positions used during the experiment are marked with purple dots.

3.4.4 Facial Electromyography

Van Boxtel [81] considers the human face as the richest source of emotional information. Thus, capture and interpretation of facial muscle activity is an important source of emotional information in media research.

According to Ravaja [57] muscle activity can be measured both via observation and with the help of facial electromyography (EMG). EMG measures the electrical signals that occur during muscle contraction. The advantage of facial EMG over observation is that EMG can also capture the hidden and more discrete movements that go unnoticed during visual observation.

Facial EMG has been compared with self-reporting by van Boxtel [81], who found that facial EMG was able to measure the emotional status with greater detail in comparison with self-reporting methods. Moreover he found that with facial EMG was possible to link certain emotions in the temporal dimension, which is impossible to do with post trial self-reporting.

The measurements are usually done by using small surface electrodes placed close to each other. The EMG is recorded from muscles known to represent certain emotions, such as sadness and happiness.

Ravaja [57] and van Boxtel [81] conclude that, the activation of the cheek muscle (*zygomaticus major*) is a good indicator of positive emotions. Nega-

tive emotions are known to activate the brow muscle (*corrugator supercilii*). Moreover, muscle activity in the muscles surrounding the eye (*orbicularis oculi*) is thought to indicate an expression of enjoyment, smile and *genuine* pleasure. Sensor placement is depicted in Figure 3.6.

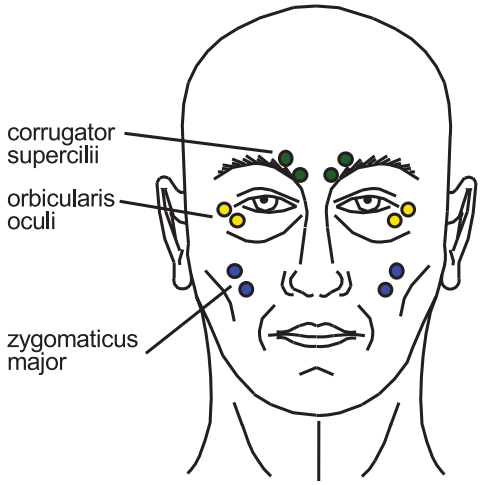


Figure 3.6: Locations for electrode placement for facial EMG measurements. For consistency, on all subjects the electrodes were fastened on the left side of the face. This is a simplified version of the figure presented in [81].

Van Boxtel[81] instructs that the following steps must be applied to the raw EMG data after amplification: high and low pass filtering must be performed to only capture the frequency range from 20 Hz to 500 Hz as this window is the known range for EMG activity. Frequencies lower then 20Hz are strongly interfered by eye blinks and movements, neighboring muscles, breathing, etc. It might also be necessary to remove the power line interference by applying a 50 Hz filtering.

Baseline EMG amplitudes and the amplitudes of the response vary between subjects. It is also noteworthy that even minor deflections in the electrode placement have a great influence on the observed signal amplitude. All these factors result in differences in affective processes, anatomical and biophysical differences. Thus it is important to process the EMG data so that it is comparable between subjects.

The obtrusive recording technique and placing of the electrodes may have an effect on the subject spontaneousness and natural behavior. Though subjects generally get used to the tactile sensors, some expressive facial movements might be affected by the presence of the sensor.

3.4.5 Orienting Response

The Orienting response (OR) is one of the central concepts of cognitive psychology and research based on psychophysical methods. Ravaja [57]

concludes that the OR is the immediate response to a change in its environment.

The OR is a consequence of any sudden, novel or meaningful stimulus triggering information processing. Characterized by a motor response and psychological changes such as a decrease in muscle activity and a phasic decrease in HR, changes in EEG (electroencephalogram) activity and a temporary pause in respiration followed by shorter and faster breath. The most notable OR as to the scope of this thesis is an increase in EDA. These responses can be detected with psychophysiological measuring instruments. As arousal increases while experiencing both negative and positive sensations, it is crucial to inspect both valence and arousal readings to find what kind of a sensation the participant experiences.

OR events are investigated using psychophysiological measurements as these events are the result of stimuli triggered by emotional effects triggered by stimuli.

3.5 Eye-tracking

In order to follow how the participants' attention was divided between the different stimuli a eye-tracking solution developed by SMI[5] was used (Figure 3.7). Their eye-tracking glasses (ETG) are worn like a normal pair of glasses. According to the manufacturer SensoMotoric Instruments (SMI) [69], "this binocular eye tracking and automatic parallax compensation delivers very reliable and accurate data over all viewing distances." This suggests that these glasses are suitable for the second screen setup.

Figure 3.7: The SMI eye-tracking glasses used for tracking of eye movements. The frames carry three cameras: two capturing the pupil movements, the third capturing the scene.

As stated by Morimoto and Mimica [45], the human eye movement can be divided into tree types of motion, smooth pursuits, vergence shifts, and

[5] SensoMotoric Instruments Ltd.

saccades. The pursuit movements keep a moving object in sight, and vergence movements keep both eyes aligned on the object of interest. Saccades are small involuntary movements made by the visual system to scan an object with the high resolution area of the eye called fovea. These saccades happen between distinct fixations. As Salvucci and Goldberg [64] state that there are several metrics to analyze, such as individual fixations, gaze durations, velocities and amplitudes of the different types of movements. Within the scope of this thesis the interest is in the number of transitions between the different stimuli and the gaze durations ratio at the two devices. This information enables objective measurements and a picture of possible effects of task switching.

All the measuring equipment is integrated into the rims of the glasses. The two cameras are facing towards the pupils capturing the eye movements at a 30 Hz sampling rate and the third high definition (HD) camera captures the scene viewed. A three point calibration is made with each participant with the software provided by the manufacturer. The wearer only needs to briefly fixate his/her eyes on the points displayed on the TV. The data is recorded and videos stored for later off-line analysis.

3.6 Participants

A media study like this requires a notable number of participants to ensure sufficient statistical validity. A target for the number of participants was set at 40 people, based on earlier research by Laine-Hernandez et al. [33] and Larsen et al. [36]. To limit the number of factors affecting the results, it was decided that both Finnish stimuli and native Finnish speakers are to be used as participants. Further normal or corrected vision and hearing was required. The test supervisor ensured that the subjects met these requirements.

The participants were recruited via ads placed around the campus bulletin boards and via online newsgroup for job ads. Furthermore, email invitations were sent to the people who had previously participated in experiments performed at the department.

A total of 43 participants (12 female, 28.6%) participated in the study. The age range was from 16 to 59 years, the average being 27 years. The participants had different backgrounds: 77 per cent had a technical one, 9 per cent were from social sciences, the rest represented other sciences. Due to technical difficulties or human-error, eye-tracking data was lost in four cases and psychophysiological measures were incomplete for one participant. The self-reporting data was good for all the participants. Each participant received two movie tickets in return for their efforts upon completion of the experimental phase.

3.7 Procedure

The experiment consisted of three sections: a pre-experiment questionnaire, the experiment itself, and a post-experiment questionnaire. The experiment session consisted of two parts: attaching and calibrating the measuring equipment followed by the actual trials.

When the participants first signed up for the test, they were sent an email greeting, containing an on-line questionnaire which the participants were asked to fill prior to arriving for the experiment. This was used to map the viewing habits and other preferences of the participants, which might explain some findings. The full set of questions can be seen in Appendix B.

At the beginning of the session, the subjects were given a brief introduction to the experiment. A rundown of the procedure was given, the duration, privacy protection, and compensation were explained. This was followed by attaching the psychophysiological sensors and putting on the eye-tracking glasses. After the calibration and base line measurements the subject was introduced to the companion application during a demonstration trial that preceded the real trial. Each individual trial followed the following scheme.

1. Stimulus
2. Reference measurement of the eye-tracking glasses
3. Post-trial questionnaire. See appendix C.

The post-trial questions were adapted to fit the trial conditions, so that no tweet related questions were asked when no tweets were present, similarly for the additional information. The two first questions were there to capture the subjective emotion. Here, two of the three SAM questions (valence and arousal) were presented. The rest of the questions were Likert-style ones on a scale from 1 (strongly disagree) to 7 (strongly agree). In some questions a binary scale was used and in some an extended scale of 1 to 9. Questions related to social presence and pleasantness were selected from the works of Chen [15], Marwick et al. [41] and Richardson and Swan [59]. Slight alterations were made to the questions to better serve the purpose of the experiment.

Once the test was over, the subject was stripped from the sensors and asked to fill in the post-test questionnaire given in Appendix D. When that was done, the compensation was handed over and the participant was free to leave.

Chapter 4

Results

This chapter will present the results found. The focus is on the most significant results in the light of the research questions. Additionally, other scope related findings are presented. The first sections start by outlining the stimulus characteristics and the effects of second screen content. This is followed by sections in which more detailed findings against the aforesaid are pointed out.

The comprehensive self-report questionnaires and the objective measurements resulted in a noteworthy data set. The large amount of data acquired gave possibilities for complex analysis methods. Nevertheless, the results presented here are mainly based on the self-report questionnaires (see Appendixes B, C and D). The findings are supported by both the psychophysiological measures and the eye-tracking data whenever possible.

Even with a big sample size, the self-reporting data was not as consistent as was expected. All self-reporting measures had a large deviation with no identifiable outliers. Outliers were defined as single data points differing more than 1.5 times the standard deviation from the mean. With statistical methods it was still possible to find significant differences in the dependent variables between some of the cases. Most analyses were performed with Analysis of Variance, a tool which is common in these types of studies. A significance level of 0.05 is used throughout the analysis.

4.1 Processing of Sensory Data

The recorded psychophysiological signals needed processing in order to be used as components in the analysis of data. First the psychophysiological data was cleaned by filtering out unwanted frequencies, followed by normalization against the baseline measures. The temporal signals were split

into epochs (a small time period) and processed using customized scripts based on Augsburg Biosignal Toolbox (AuBT)[1]. For each epoch, an average value was calculated allowing a reasoned and practical analysis against other signal data [31].

As stated previously, eye tracking glasses were used to measure the distribution of attention. The recorded gaze fixations were expressed by number and categorized into three: main screen, second screen and other. The category labeled "other" was used for fixations that did not fit into the two main categories or for erroneous or incorrectly identified fixations. A semi-automatic annotation tool was used to classify the fixations.

Heart rate data was successfully recorded but some trouble arose when analysing it in a way which enabled that the data to be used as a measure for arousal. This challenge was mainly due to lack of documentation and software needed to process the data. In addition, respiratory data would have been useful to accurately separate the effects of the SNS and the PNS.

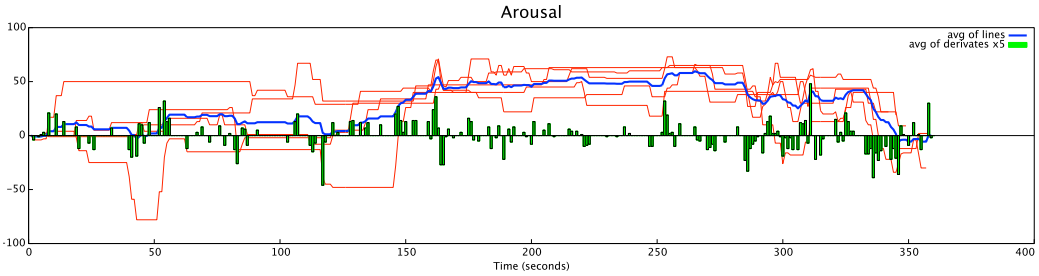
4.2 Video Sentiment

In order to get baseline values for video valence and arousal, the results of the ground truth collection are now examined. The eight video stimuli used in the trials were validated by five people, present and former members of our research group. Each validator completed the validation over a period of a few days. No separate session was organized, as the validation could be done online among other tasks.

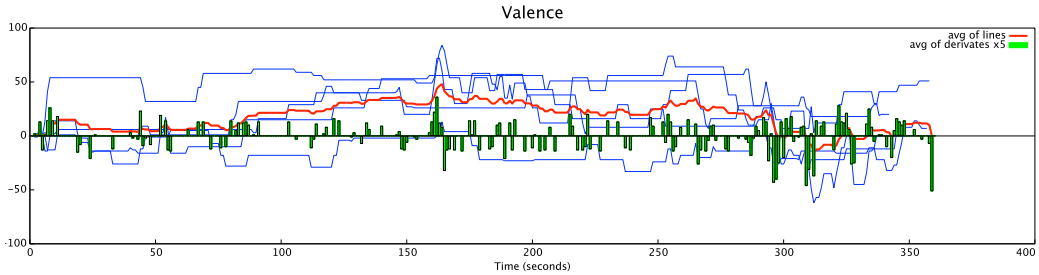
Both a statistical and a visual inspection of the results indicate that the clips seem to evoke the same feeling across all the validators. The results of stimulus `robbe` was chosen to demonstrate this in Figure 4.1. The magnitudes of feelings differ but if measured on two levels (positive and negative) they match all over and the emotional changes seem to take place almost simultaneously. To visualize this, the derivate is plotted as bar charts in the figures.

The mean value of the individual ratings was calculated. The changes in mood correlated positively with the measured psychophysiological values for the positive derivate. In other words, when validation data suggested a positive change, the same positive reaction could be found in the psychophysiological measures as well. The same effect was not as strongly present in the negative feelings, as no negative psychophysiological response was captured. Only lack of positive emotion was found. This was

[1] For more details see webpage at `http://www.informatik.uni-augsburg.de/en/chairs/hcm/projects/tools/aubt/`



(a) Arousal values of `robbe`. Red line represents individual validations. Blue is the mean and green bars visualize the change in the mean.



(b) Valence values of `robbe`. Blue lines are individual validations, red line represents the mean. Green bars visualize the change in the mean value.

Figure 4.1: Visualization over the affective ground truth values gathered for `robbe`.

probably so, because the feeling returned to a "neutral" state unevenly between the subjects.

When looking at emotional data captured with SAM in the post-trial questionnaire, the same correlation could not be found, not with either psychophysiological data nor with the manual validation data. This suggests that the participants had a difficulty to perceive the affective state after the stimulus had been shown or that the personal preferences (e.g. some like sports when others hate it) explain why the emotional measures were so scattered. It may also be that the participants were asked to reflect the whole viewing experience, not just the TV program, while filling in the SAM values.

The above findings and the lack of robust validation values (a too small group of validators) led to the use of test condition I (only TV stimulus) as the ground truth for analysis of the three other test conditions.

To conclude, temporal self-reporting shows promising results. A bigger group of validators and research with a scope solely focusing on this topic could result in more accurate findings. Temporal validation might also give more emphasis to the beginning of the stimulus, as viewers tend to emphasise the end more when scoring. This is known as the *recency effect* [55]. Further insight into this issue is outside the scope of this thesis.

4.3 TV Content Characteristics

In order to be able to discuss the effects of second screen content an understanding of the baseline characteristics between the different stimuli is needed. A subset of questions in the post-trial questionnaire listed under the section *TV-program* are here used to describe the characteristics of the TV stimuli (Appendix C).

To compare the interestingness of the four genres *when no second screen content was shown*, plotted in Figure 4.2, a one-way ANOVA was conducted. A significant effect in TV interestingness was found between the genres $(F(3, 82) = 5.12, p = .003)$. Further post-hoc comparisons using Tukey HSD revealed that the mean score for magazine was significantly different from the rest $(M = 5.55, SD = 1.18)$. However, the other conditions, documentary $(M = 3.95, SD = 1.91)$, reality $(M = 3.95, SD = 1.59)$ and sports $(M = 4.00, SD = 1.75)$, did not significantly differ from each other.

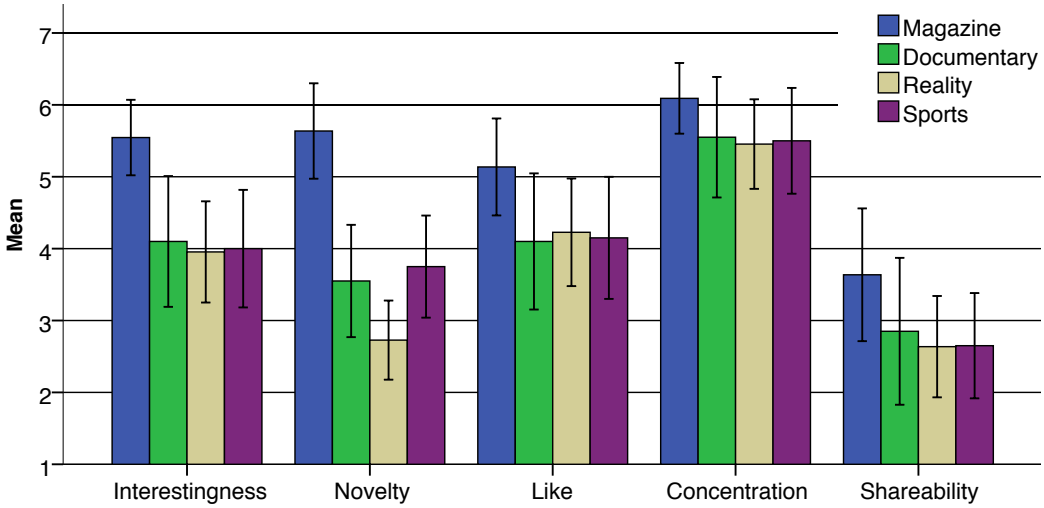


Figure 4.2: Visualization of self-reported genre interestingness, novelty, like, concentration and shareability with 95% confidence intervals.

In addition, it is noteworthy to recognize that content interestingness correlates with how much the participants liked the content. A Pearson correlation coefficient was computed to assess the relationship between these. There was a notable positive correlation between the two variables within test condition I, $r = 0.87, n = 86, p < 0.001$.

Similarly, program novelty was analysed comparing the genres $(F(3, 82) = 14.46, p < 0.001)$. Tukey HSD shows that magazine genre differed from the rest $(M = 5.64, SD = 1.50)$. It provided more novel content than docu-

mentary ($M = 3.55, SD = 1.77$), reality ($M = 2.75, SD = 1.25$) and sports ($M = 3.75, SD = 1.52$).

Further, content genre did not have an effect on how much participants concentrated on the TV content ($F(3, 82) = 0.85, p = .47$). With no other stimuli to compete for the attention, the differences remained non-significant.

Neither were significant differences found between the genres in the tendency to share, even though magazine programs got a slightly higher score compared to the other genres ($F(3, 82) = 1.45, p = .23$). In spite of this it is important to notice that the overall sharing score is very low, very few seemed to be ready to share the program.

Assessing the validity within genre. In addition, the same dependent variables are plotted by individual programs in Figure 4.3. Here it can noted that within genres there are some differences between individual programs, but statistically significant differences were only found within the magazine genre.

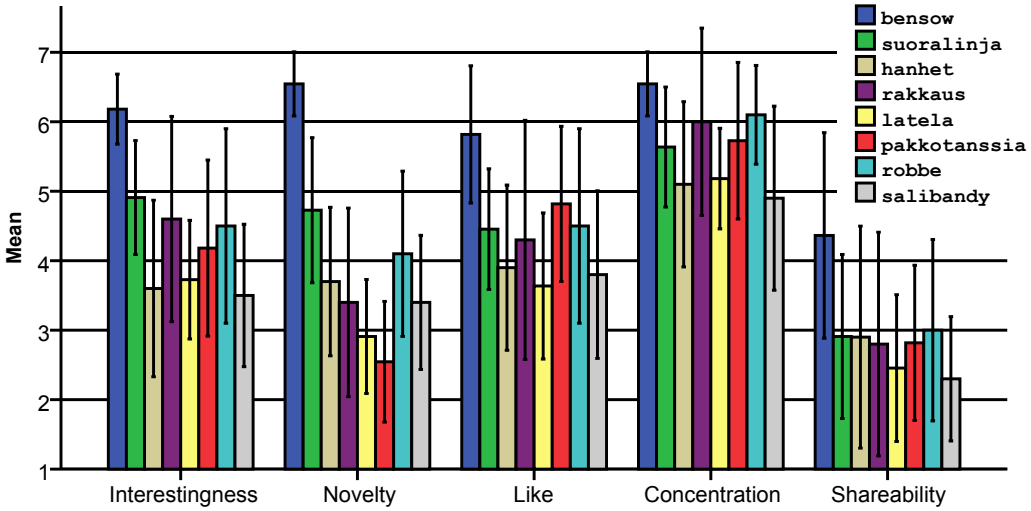


Figure 4.3: Visualization of self-reported TV content interestingness, novelty, like, concentration and shareability with 95% confidence intervals.

Performing an independent-samples t-test reveals that `bensow`'s interestingness statistically differs from `suoralinja`'s interestingness ($t(20) = 2.95, p = .008$). Similarly, content novelty differs in favour of `bensow`'s over `suoralinja`'s ($t(20) = 1.82, p = .002$). Further, participants liked `bensow` more compared to `suoralinja` ($t(20) = 2.31, p = .032$). A t-test did not detect a difference in `suoralinja` and `bensow` shareability ($t(20) = 1.71, p = .102$). The stimulus genres can be considered having been well chosen.

Looking at the plotted facial EMG data (Figure 4.6) similar behaviour can be seen within genres regarding the effects of tweet presence. In all of the cases, the presence of tweets affected the EMG in the same way. If a change in affective values could be seen in one of the stimuli a change in the same direction could be seen in the other stimuli as well within the same genre. Additionally, EMG responses seem to be of the same magnitude within a genre. These findings have statistical support as significant differences between genres could be found in both conditions I and II.

4.4 Enriching the TV Content with Tweets

In this section, viewing conditions I (TV only) and III (TV+tweets) are compared. In other words, it is studied how the introduction of social content on the second screen affects the otherwise passive viewing experience. The self-reported TV measurements are compared with the two conditions, and the results are presented in Figure 4.4.

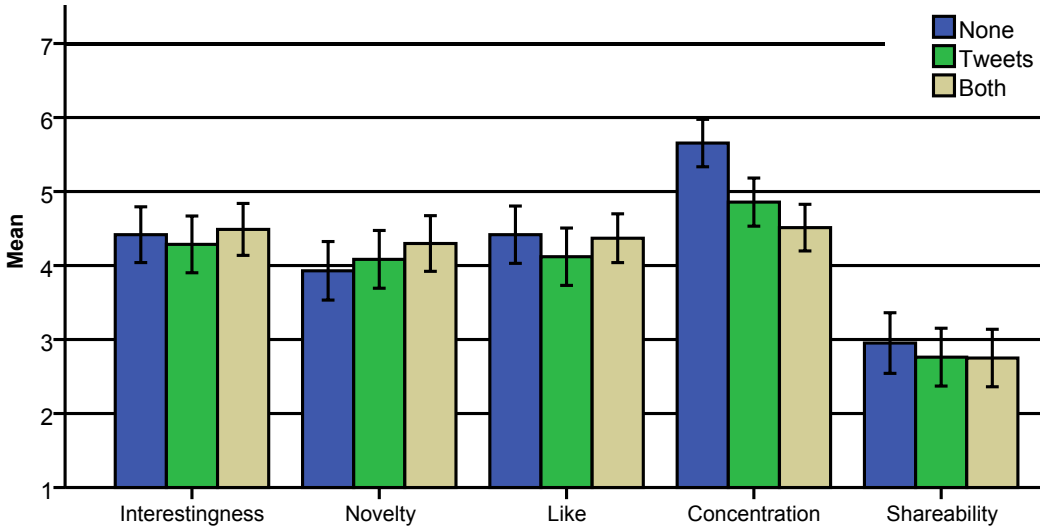


Figure 4.4: Comparison of effects on TV content in conditions *no tweets present* and *tweets present*.

When inspecting Figure 4.4 it is clear that the introduction of tweets did not have an effect on how the participants experienced the majority of the variables measuring TV content features. This suggests that the different media do not merge into one, but viewers still consider the different media as separate.

As expected, a decrease in TV concentration can be seen upon tweet introduction. A t-test confirms a significant decrease of $.81$ when tweets are shown $(t(170) = 3.61, p < .001)$. This finding is in line with the multitasking literature and the studies presented in Section 2.1.

Further, when adding information to go along with tweets (condition IV), similar results can be seen: no significant effect on the TV content related measurements, except for the further decrease in TV concentration. This indicates that the participants still distinguished the two physical media from each other. In other words, second screen content cannot improve the TV content itself, but it might improve the overall viewing session experience.

Analysing the ETG-data, in Figure 4.5, a clear attention shift towards the tablet device can be seen, assuming that when only TV content was present the attention is almost fully (95%+) directed towards the TV. Similar error levels (plotted `Other`) can be seen as in the other conditions. When the tweets were introduced, the focus on TV dropped by a third, in other words the tablet received $1/3$ of the attention. No statistically significant differences could be found between the different TV stimuli in the distribution of visual attention.

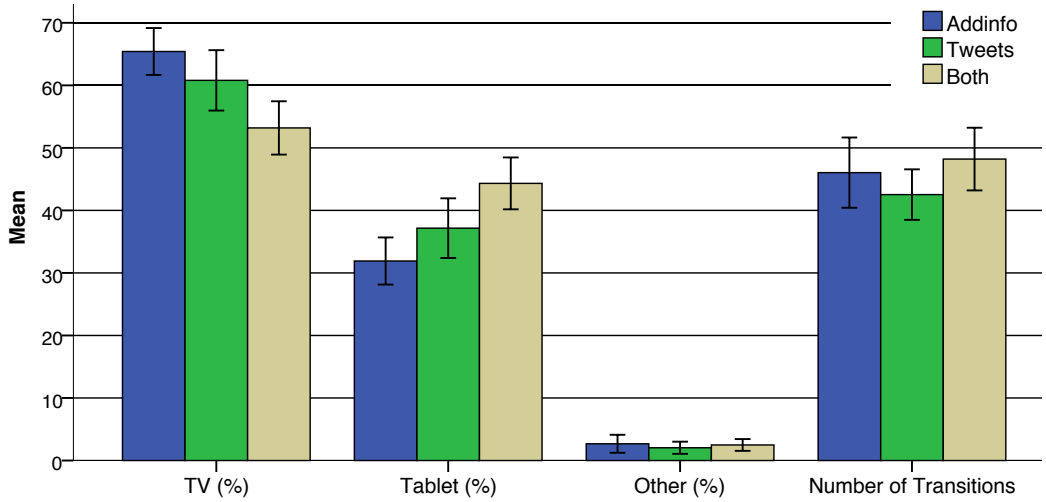


Figure 4.5: The distribution of visual attention in the different stimuli conditions. Bars represent a ratio of attention between the stimuli. In the plain TV condition, it is assumed that nearly all (>95%) of the attention is directed on the TV screen.

Based on Figure 4.5 and later findings presented in Section 4.6 it can be concluded that in situations where TV content was considered boring,

the presence of tweets shifted the attention more towards the tablet when compared to situations where the TV content was considered interesting. This can be interpreted to mean that even boring tweets were worth reading during a boring program, even though the reported arousal levels did not increase, as reported in Section 4.4.5.

4.4.1 Tweet Sentiment

The SentiStrength algorithm, discussed Section 3.2.3, was used to predict the sentiment values of the tweets. Upon later inspection, the results of the tool were dissatisfying as many of the tweets were poorly assessed. A comparison between the results of manual assessment and the job done by SentiStrength revealed weak positive correlation between the two ($r = .38, p < .001, n = 105$). A common random sample of 105 tweets was used for manual validation made by two employees. The results of these two human annotations correlated somewhat better ($r = .43, p < .001, n = 105$). With the resources available at the moment, it was agreed that no further manual scoring of the tweets would be done as the fact remains that no trustworthy sentiment data could have been achieved. Based on these conditions it was decided that tweet sentiment basted result analysis would be dismissed this time.

Nevertheless, Table 4.1 presents the cumulative sentiment score for the tweets for each stimulus. The cumulative value is simply the sum of the negative and positive scores for all the tweets associated with the program. In the same table the ratio of individual positive and negative tweets are presented. For a comparison also the participant reported tweet like is presented.

Table 4.1: Results of tweet sentiment analysis done with SentiStrength algorithm. Also the reported tweet like is listed.

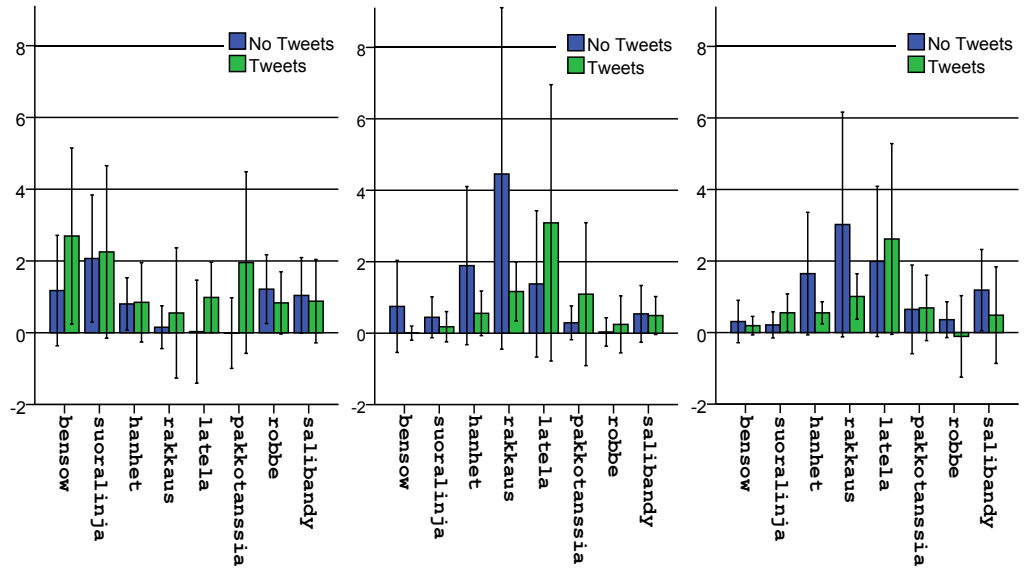
	bensow	suoralinja	hanhet	rakkaus	latela	pakkotanssia	salibandy	robbe
SentiStrength	4	0	5	95	17	45	8	17
Positive (%)	76.5	18.9	31.4	64.8	39.4	49.2	21.0	41.3
Negative (%)	23.5	81.9	68.6	35.2	60.6	50.8	79.0	58.7
Reported tweet like	3.50	3.20	2.00	2.00	3.00	2.27	2.55	3.35

In Figure 4.6 the EMG data is plotted in conditions I (TV only) and II

(TV + tweets) separately for each stimulus. Each bar represents the mean response to the stimuli compared with the baseline mean measured before the experiment started. Again the confidence intervals are quite big, so no statistically significant findings can be stated.

Nevertheless, if in the light of the figures, it seems like the negative response is increased exceeding the no-tweets condition except for the both sports stimuli. Similarly, positive response decreases for the most part in both ZM and OO. Only `latela` seems to make an exception showing more positive response (ZM) when the tweets are present but still lacking the indication of true emotion (OO).

It can be argued that the reason for this is not necessary found in the tweets them selves but is more likely to be a result of the human behavioural aspect of wrinkling the eyebrows during visual focusing, during tasks such as reading.



(a) EMG CS (-) (b) EMG ZM (+) (c) EMG OO (++)

Figure 4.6: Results of facial EMG in conditions I and II (μV).

Further findings and discussion on EMG signals and tweet presence can be read in Section 4.4.5 and more simple EMG plots can be seen in Figure 4.10 in the same section.

4.4.2 Tweet Characteristics

In this experiment a design decision was to use real tweets written by real people with all the video stimulus. After each trial, the participants were asked different tweet related questions this sections presents the main characteristics found in these answers.

Briefly said, the tweets were not that well received. Most of the comments we heard were that tweets were improper: they did not come in real-time, they were not closely related to the program, the tweet flow was too vivid, i.e. there were too many tweets, they drew too much attention from other media giving no real value to the experience.

Apart from that, it was found that the more interesting the tweets were, the more participants liked them ($r = .87, n = 172, p < .001$). The added value the tweets brought also followed the reported interestingness ($r = .77, n = 172, p < .001$). All tweet related measurements are plotted in Figure 4.14. Thus, it is exceedingly important to pay attention to the method of tweet selection for a desired increase in media viewing experience, for example, only to bring retweeted tweets to the second screen, as retweeted tweets most likely are interesting in some way.

In order to better visualize the differences, the four genres are regrouped into two. In Figure 4.7, it can be seen that all the tweet measurements score higher in the entertainment genre in comparison with when displayed with more informative programs. Note especially the higher increased added value and the experience brought by tweets. This would indicate that tweets are more welcome along TV viewing when watching entertainment than when watching e.g. documentaries. Additionally the tablet was found less disturbing while watching programs with lighter content.

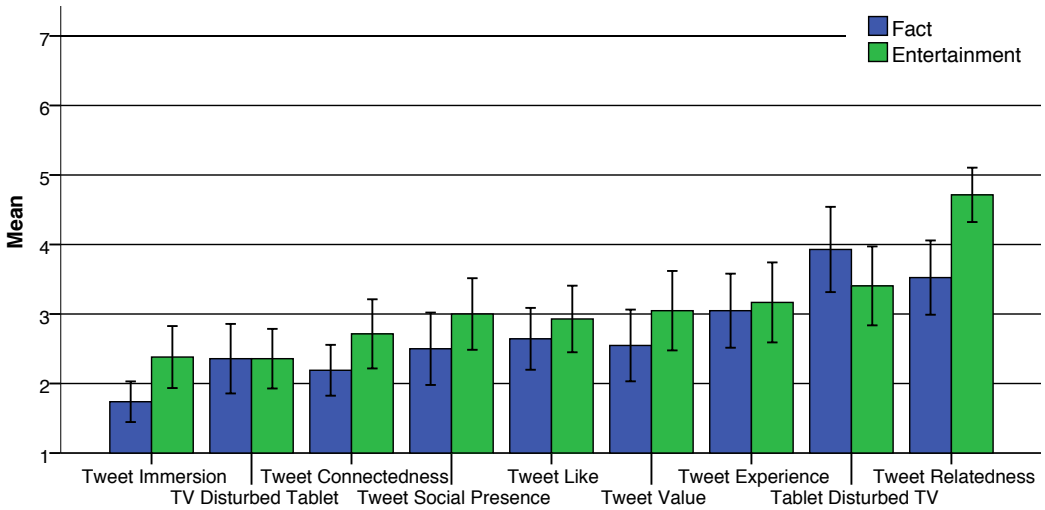


Figure 4.7: A comparison between tweet measurements in categories fact (magazine and documentary) and entertainment (reality and sports).

In the post experiment questionnaire it is easy to see how the tweets divided opinions (Appendix D). Some praised the tweets as giving something to do during a dull program, others gave more negative feedback describing the tweets with phrases like: "Tweets were completely unnecessary", "I felt like reading old messages from a night time TV chat", "The tweets felt a bit strange" or "Tweets irritated me".

4.4.3 The Number of Tweets

When planning the experiment, an educated guess was made regarding the number of tweets aiming at a pleasant tweet experience. This estimation had to be done, as even the most frequently tweeted program only received very few tweets during the broadcast.

The following results reveal that the guess turned out to be quite unsuccessful as most people found that there were too many tweets seen in elevated perceived tweet count. This observation can be made, especially as the tweets in most cases were not that closely related to the TV content. Contradictory still, multitasking levels remain manageable as will be seen next.

In test condition I (TV only) no multitasking effort is expected. As seen in Table 4.2, the reported multitasking effort rose across all stimuli upon tweet introduction. Still the participants consider the multitasking effort manageable. As the score remained below 4, the value considered neutral. Out of interest and for comparison, multitasking and tweet scores for viewing condition IV (both tweets and additional info) is also reported in the table.

In spite of the above, the participants felt that the tweet stream was too vivid as all values (except for stimuli `hanhet`) exceed the value considered neutral on the self-report questionnaire. As the number of tweets was not varied across the experiment, it is difficult to say what the optimal number of tweets would be. It can be seen that further increasing the number of second screen stimuli resulted in higher tweet count scores stating that there were too many tweets shown.

The polychronicity score did not explain the reported multitasking effort in the case where only tweets were shown ($r = -.03, n = 80, p = .83$). The same is true for the case were both tweets and additional info were shown ($r = -.20, n = 80, p = .730$). It would be expected that a person with a higher polychronicity score would tolerate a more intensive media environment with the same reported levels of multitasking effort as a less polychronic person. In the same fashion, the tweet count was compared against eye-tracking data, but there are no significant findings

Table 4.2: Tweet count, tweet feed speed, reported multitasking effort and perceived tweet count in conditions where *only tweets* and when *both tweets and additional info* were shown.

	bensow	suoralinja	hanhet	rakkaus	latela	pakkotanssia	robbe	salibandy
Number of tweets	34	53	35	54	33	63	63	62
Tweets per minute	6.92	10.50	7.58	9.97	6.58	13.60	10.56	11.89
MT Effort (t)	2.80	3.50	2.73	3.18	2.09	3.09	3.00	3.09
MT Effort (b)	3.73	3.64	2.45	3.45	2.80	2.40	4.00	3.91
Tweet Count (t)	4.40	5.10	3.40	4.55	4.55	4.91	5.00	5.18
Tweet Count (b)	5.09	5.55	5.09	5.64	4.50	5.10	4.09	4.91

t = tweets
b = both tweets and additional info

to report for either the number of transitions made between the devices ($r = -.01, n = 80, p = .952$) or the time spent focused on the tablet ($r = -.01, n = 80, p = .941$).

4.4.4 Tweet Interestingness

The participants systematically noticed how well the tweets were related to the TV content. As stated in Chapter 3, some tweets more closely related to the TV content than others. In a between stimuli ANOVA analysis of self reported tweet interestingness ($F(7, 164) = 3.33, p = .002$) and tweet relatedness ($F(7, 164) = 15.78, p < .001$) a noticeable difference could be found between the different TV contents. E.g. stimuli `hanhet` and `rakkaus` had less related tweets which could be confirmed by the self reporting data. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the `rakkaus` ($M = 2.36, SD = 1.22$) and `hanhet` ($M = 2.41, SD = 1.217$) tweets significantly differed from the most related tweets `robbe` ($M = 4.95, SD = 1.29$) and `suoralinja` ($M = 4.86, SD = 1.32$) which both, we and the participants, considered related to the program. These differences can be seen in Figure 4.8.

Reported tweet interestingness correlated highly with how much they liked the tweets ($r(172) = .86, p < .001$). In addition, tweet interestingness had a moderate correlation with the added value brought by tweets

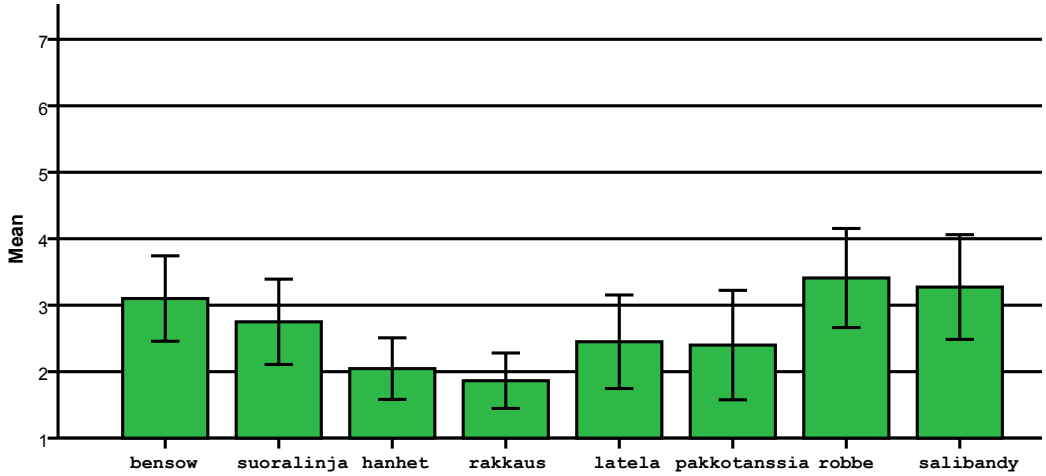


Figure 4.8: Participant-reported tweet interestingness by TV stimulus.

$(r(172) = .77, p < .001)$. Furthermore, when the mean tweets relatedness correlated strongly with how much the tweets were liked $(r(8) = .85, p = .007)$. This would suggest that it is important to choose tweets that are related to the TV content. This makes the tweets interesting and further liked, which results in an improved viewing experience.

A link between reported likelihood to share the TV content with friends and actual retweets was not found. This may be the result of overall low share-score or due to the fact that the individual tweets did not represent the content in a desired fashion, true to the self-representation tendency of our participants.

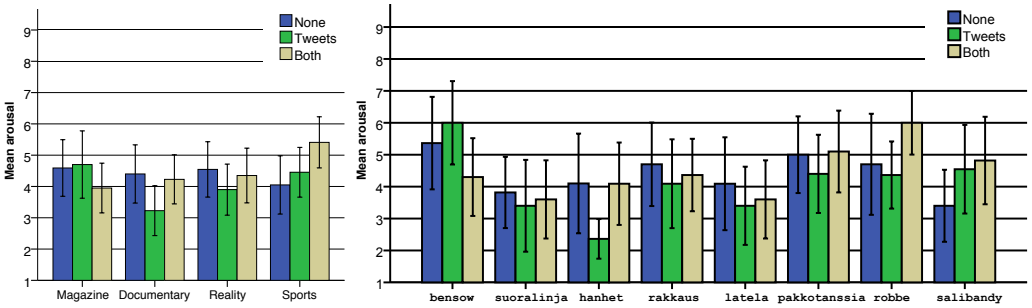
4.4.5 Tweet Effect on Valence and Arousal

As discussed in Chapter 2, arousal and valence are commonly considered essential measures in subjective viewing studies. It could me assumed that an increase in second screen stimulus would result in an increase in arousal and valence: plain TV would have the lowest values followed by either tweets or additional information with somewhat similar readings. A simultaneous display of all the stimuli would be the most intense. In the end, this turned out to be rarely the case.

An independent-samples t-test was conducted to compare valence in *tweets present* and *no tweets present* conditions. No significant difference was found between tweets $(M = 5.19, SD = 1.75)$ and no tweets $(M = 5.19, SD = 1.68)$ conditions; $t(342) = 1.10, p = .282$. Similarly for arousal, no statistical differences were found $(t(342) = .67, p = .506)$ between tweets

being present ($M = 4.22, SD = 1.94$) and no tweets present ($M = 4.35, SD = 1.95$).

Visual inspection of relatively big error bars in Figures 4.9a and 4.9b reveals the fact that the subjects did not agree on the affective values. This was contrary to the expectations. Analysis of standard deviations and kurtosis also confirms this. No outliers could be identified using visual or statistical methods. The wide spread of answers was consistent throughout almost all self-report questions in most of the stimulus combinations discussed in Chapter 3.



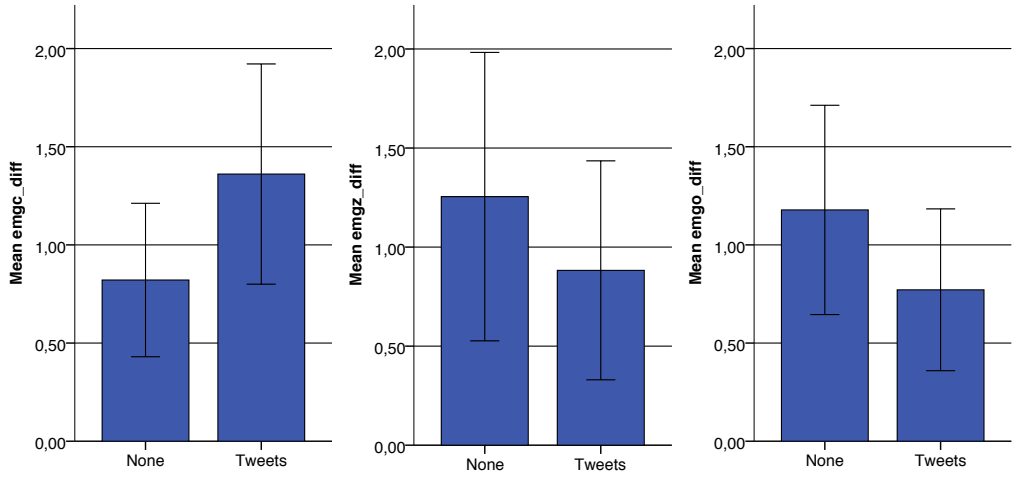
(a) Arousal by genre

(b) Arousal by TV content

Figure 4.9: Visualization of self-reported arousal in different second screen conditions, 95% confidence intervals.

To support the self-reported data, SCR count was used here as a measure of arousal, the baseline reading was subtracted from each trial SCR measurement to get the readout from the stimuli effect. Measurements based on the SCR count are in line with the above, no statistically significant finding could be found between conditions I and II, neither in the overall score or within stimuli condition.

Looking at Figure 4.10 it is clear that the introduction of tweets decreased positive facial EMG response (*zygomaticus major* and *orbicularis oculi*) and increased the negative response (*corrugator supercilii*). In the figure the baseline measurements are subtracted from the trial measurements, which gives us the objective reading on the emotional state of the participants. Although a difference in EMG response can be seen in the figures, no statistical differences could be found between these two groups. Regardless of the above it is encouraging to see some, yet very weak, correlation between self reported valence and the facial EMG data. The Pearson product-moment correlation coefficient, between self-reported valence and the EMG data were as follows: EMG-O ($r = .33, p < .001, n = 186$), EMG-Z ($r = .25, p < .001, n = 168$) and EMG-C ($r = -.11, p = .141, n = 168$).



(a) EMG CS (-) (b) EMG ZM (+) (c) EMG OO (++)

Figure 4.10: Results of facial EMG in conditions I and II for each TV stimuli used (μV).

Looking at the data presented in this section, it seems that tweets would just affect the experience negatively. Especially EMG measurements clearly indicate a decrease in the experience as they indicate an increase in negative response and a decrease in the positive. But it can be argued that this is a natural response to increased focusing as the result of reading the second screen content. Thus it could be argued that self-reported data should be given more weight when interpreting different conditions.

4.5 Social Media Findings

The pre-experiment questionnaire (Appendix B) was used to map social media behaviour. The findings revealed that our participant population was not that familiar with Twitter and that the activities done on social media benefit more from the features and nature of Facebook. The use of social media was very common. Only one participant reported that he/she is not part of any social media.

Participants reported that the activities performed on social media try to mimic real life activities, content is created with certain friends in mind and topics discussed follow real life interests and social values. This is the same way you would choose your social environment in real life.

When asked what social presence means to the participants, a common theme emerges: being able to communicate with people, whether it be via a communication device or face to face. These two situations were frequently

reported and strictly separated form each other. In a situation when face-to-face interactions happen the participants prefer to avoid device use and fully concentrate on the situation at hand. On the other hand, while consuming social media the participants usually report using it for chatting and taking part in conversations. Whatever the case might be, active participation is required of all parties.

What comes to social media, active participation is considered important, even though the level of involvement is reported in a more shallow way, "likes and commenting" was a common answer. A majority of the participants report social media as a tool for picking up new ideas for events and topics for conversation and using the chat feature to plan face to face interaction. Some even state that it is a place to spend a good time for picking up funny stuff. With TV content in mind, this is an ideal place for viewers to pick up new programs to watch.

4.5.1 Social Media Consumption Habits

The background questionnaire was used to find out social activities and habits on the social media of choice. The participants were asked to choose which social media they consider using, multiple medias could be chosen. Facebook was clearly the most popular, with 90 per cent of participants using it. The second most popular social media was LinkedIn, a work- or business-oriented social media site, with 36% participants. Twitter came as the third most used media with a 27% user base. These choices were not excluding, so one could select several medias.

The reported use of Facebook is in line with the activities done in social media. When asked what relationships people maintain via social media all reported that they keep in touch with friends. In addition, nearly 70% reported being in contact with different societies, nearly 70% reported keeping in touch with different social circles. Finally one third reported using social media to maintain contacts with family and loved ones. As said, these actions can be considered quite private. The public nature of Twitter is not very suitable for such activities.

A question about Twitter use revealed that the participants did not feel like being part of the Twitter community, the answers were really surprising with the common second screen practices in mind. 63% reported that they did strongly disagree, 18% of the participants said they disagreed, 12% answered neutral and only 7% agreed. None of the participants did strongly agree.

Participants were asked to choose the three most important activities performed on social media, the results can be seen in Figure 4.11. When

asked what participants considered as the most important activity on social media, 55% reported the ability to follow what others were doing. 45% reported that chatting with individual persons is the most important activity.

When asked about the second most important thing the opinions were separated even more: two out of five (40%) reported individual chatting, almost an equal amount (35%) reported following other people's activities. The remaining 25% were divided equally between sharing their own content and commenting other people's activities.

The majority (45%) of the participants reported commenting on other's activities as the third most important activity. Even now sharing one's own activity did not reach the majority of answers (43%). The remaining share was equally divided between individual chatting and following others.

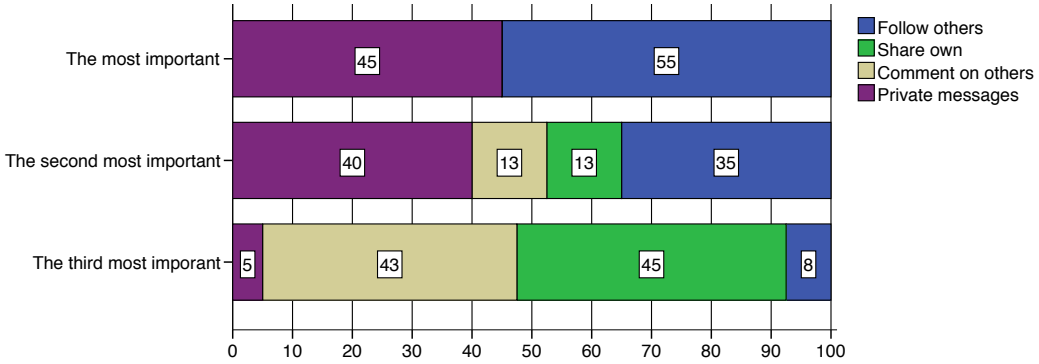


Figure 4.11: The most important activities performed on social media.

4.5.2 Sharing and Content Creation

The participants had a tendency to create social content particularly with their friends in mind. Half of the participants (50%) told that they recognize this behaviour, 23% did not have an opinion and 27% did not find this behaviour in their social media habits.

The above findings relate to the findings regarding content creation. 73% reported that they do not create content with strangers in mind. 15% do not really think about whom they direct their posts to, and 12% agree at some level of thinking about strange people as well.

What comes to the content itself, opinions vary. When asked if the participants care about what they put up on social media content vice, the answers are close to normally distributed around the neutral answer choice ($M = 3.13, SD = .99$), slightly skewed towards the caring answer.

This would suggest that all TV content could be up for sharing if considered interesting enough or worth sharing by another meter. It could also be that the sharing behaviour is a result of impulsive behaviour, stumbling upon the right kind of content, or it does not matter what it is about.

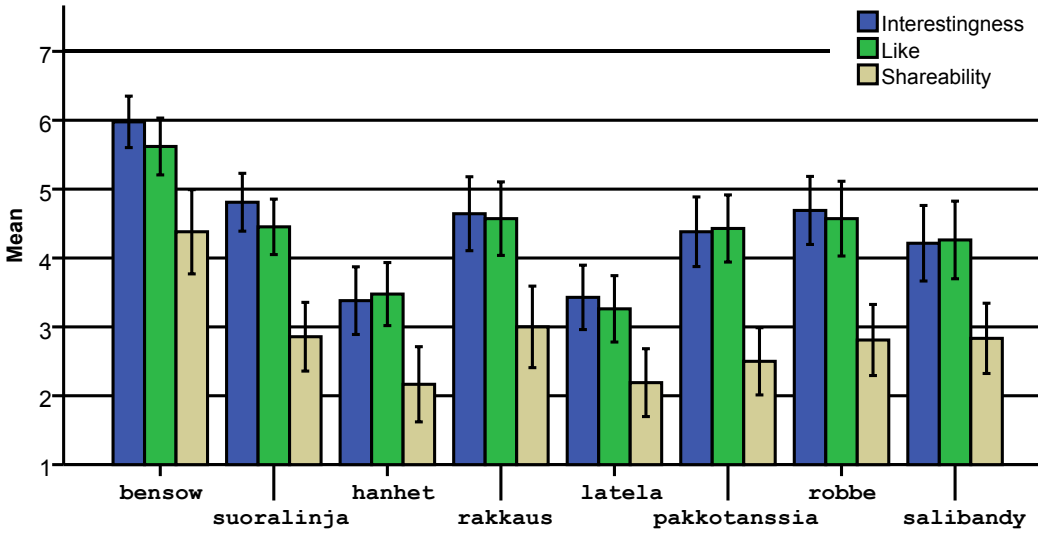


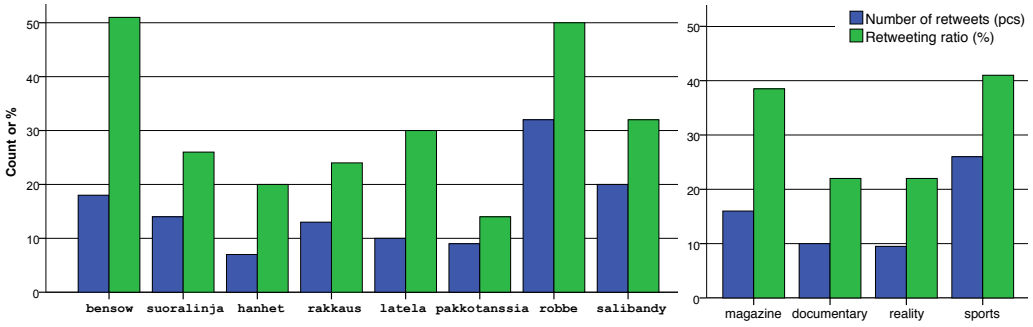
Figure 4.12: Tendency to share, TV interestingness and like scores plotted. With 95% confidence intervals marked.

Data from the post trial reports revile that good (interesting and liked) television broadcast increases the urge to share, as shown in Figure 4.12. Reported likelihood of sharing the TV content correlates moderately with both TV interestingness ($r = .61, n = 84, p < .001$) and TV likability ($r = .59, n = 84, p < .001$). No significant differences in TV shareability was found between the different TV stimuli ($F(7, 76) = 0.64, p = .725$) or the different genres ($F(3, 80) = .71, p = .550$). These were the self-reported values in Section 4.5.3, which show how the participants actually responded.

4.5.3 Retweeting Behaviour

In total the experiment had 437 unique tweets to display with the eight TV stimuli. Out of the 43 participants 26 chose to retweet once or more that is 60 per cent of the sample group. More detailed information on the tweets can be seen in Table 4.2. Retweets were used as a measure of involvement in social media, thereby indirectly measuring social presence and immersion in addition to second screen involvement. Thus, each retweet a participant made was recorded. There were prominent differences in the number of

retweets between genres and individual programs, as seen in Figure 4.13a and Figure 4.13b. The sports genre gathered a lot of retweets both in absolute value and in relation to the total tweets presented. In contrast, the magazine genre got only half as many retweets in sheer number, but when looking at the relative number, it received as many retweets as magazine. These differences raise the query what provoked the differing retweeting behaviour.



(a) Retweets by program (b) . . . and by genre

Figure 4.13: Relative and absolute numbers of retweets plotted.

Comparing the occurred retweets with the subjective measure of tweet relatedness, there seems to be some support to explain why magazine and sports programs got relatively more retweets compared with the other two genres. It seems as if the former performed better on all the self reported tweet measures in comparison with the latter. This is visualized in Figure 4.14.

Looking at the figure, one measure stands clearly out: it seems like people would have read sports related tweets significantly more often compared to other genres regardless of the tablet application (Read Tweets). Yet, it is interesting to see that the magazine genre has triggered so many retweets. If comparing the tendency to read tweets with the occurred retweets, the difference is manifold. Hence, in the magazine genre the added value brought by the second screen application is self evident.

Another feature that stands out is that the tweets in the documentary genre do not relate with the program at all, which might explain the low retweet rate. Furthermore, the documentary tweets performed worse on all counts in comparison with the other genres.

Noteworthy is that in the TV stimulus there were two programs that were strongly scripted around audience tweets (`suoralinja` and `pakkotanssia`). However, with a magnitude of tweets, only a few of them resulted in retweets

Table 4.3: Summary of tweets, retweets and the retweeting ratio in each TV content.

	bensow	suoralinja	hanhet	rakkaus	latela	pakkotanssia	robbe	salibandy
Stimuli tweet count	32	53	35	54	33	63	63	62
Total retweet count [1]	18	14	7	13	10	9	32	20
Unique retweet count [2]	17	13	6	11	10	9	24	10
Retweet ratio (%) [3]	53	26	17	20	30	14	38	16
Retweets/participant [4]	.40	.30	.14	.26	.23	.21	.56	.23

1) Cumulative count containing all retweets by all participants
2) Multiple retweets for a single tweet are only counted once
3) How many of the tweets got retweeted
4) Average retweets per participant, if all 43 participants would have retweeted

in comparison with programs where no scripted Twitter involvement was done by the broadcasting company.

Comparing polychronicity, the BIS and BAS scores between those *who retweeted* and those *who did not retweet*, no significant differences were found in BAS or polychronicity scores. Yet it is worth noteing that the polychronicity index was higher among those who tweeted ($M = 40.92, SD = 7.78$) in comparison with those who chose not to retweet ($M = 36.31, SD = 9,09$). This difference might have been significant with more participants.

On the other hand, there was a significant difference in BIS scores between the ones who retweeted ($M = 18.00, SD = 2.78$) and the ones who did not ($M = 20.31, SD = 3.05$); $t(39) = 2.50, p = .017$. This would indicate that people that are less afraid of the possible negative consequences are also more often ready to retweet.

To conclude, the overall low results in Twitter measures seen in Figure 4.14 are in line with the reported low Twitter usage reported figures obtained by the background questionnaire. Yet some surprising results were found, such as when present, the second screen application encouraged people to use Twitter more actively. How well the tweets related to the program and the program genre had a significant difference in participant tweet involvement.

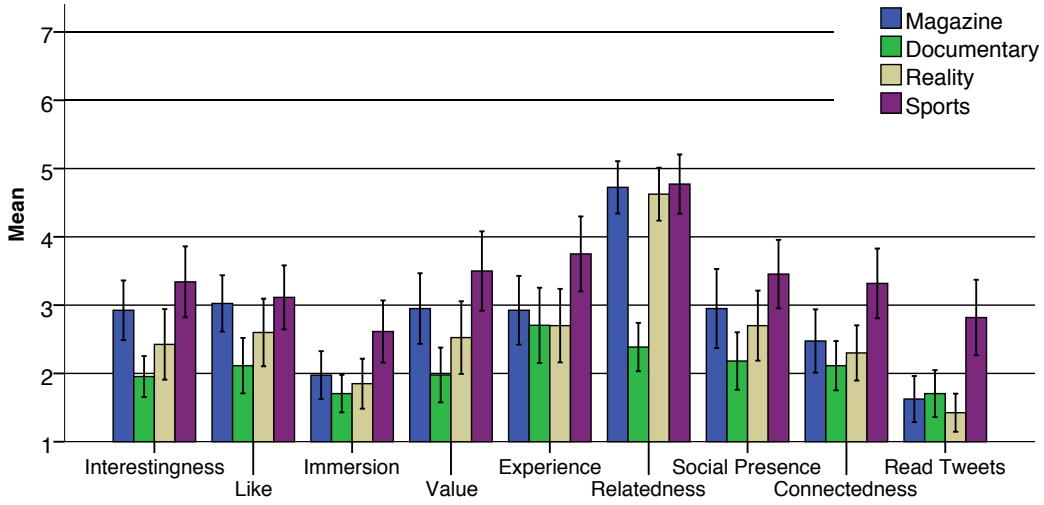


Figure 4.14: Visual representation of the tweet-related self-report questions.

4.6 Multi-Screen Behaviour

In multitasking situations, as in the experiment of this work, literature usually discusses the effects of multitasking on performance and increased effort. These often feature memory tests and tasks to be performed, but in this experiment no such tasks where given. In this work a different approach is taken by examining the effects of second screen presence on the different measures in the context of the viewing experience. In this thesis, the *multitasking measures* are based on data collected by the questions presented under the section *multitasking* in the post-trial questionnaire (Appendix C).

In Figure 4.15 multitasking measures are plotted. In most of the multitasking metrics no significant differences were found between the different TV stimulus conditions. This did not apply when participants were asked which screen the participants considered the main media ($F(7, 336) = 2.602$, $p = .013$) and if the participants would had searched for additional info without the companion application ($F(7, 336) = 8.374, p < .001$). To summarize, in less interesting and slow paced contents (`hanhet` and `latela`) the second screen is considered more often as the main media seen as slight elevations in Figure 4.15. Further, voluntary search of additional info is more likely to take place while watching magazine and sports.

Now, comparing the reported interference of TV and tablet device. Although there is no statistical support, it seems that the tablet is considered

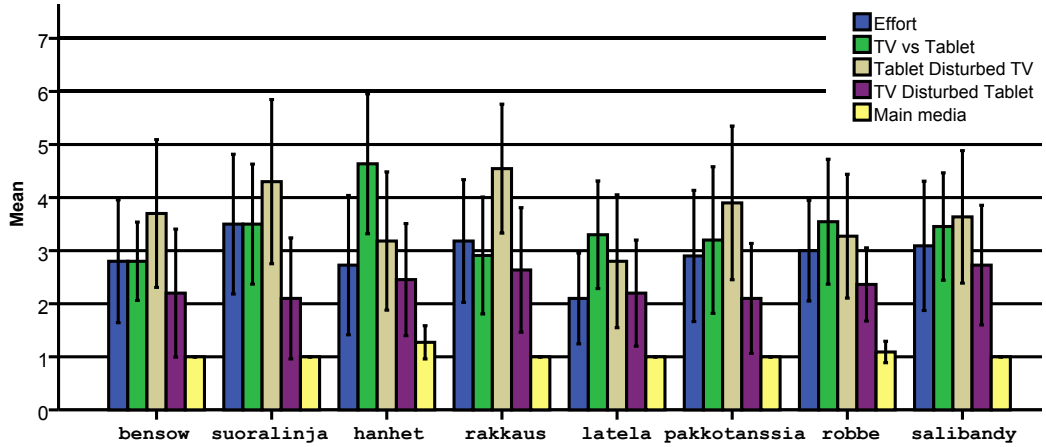


Figure 4.15: Reported multitasking measures when tweets are shown. *Main media* ranges from 1 (TV) to 2 (tablet). The others range from 1 to 7.

less disturbing in cases where there are tendencies to regard the tablets as the main media as seen in the case of documentary program `hanhet`.

The subjective progression of time might in some situation get distorted [63]. Yet, when analysing the eye tracking data it seems that media consumers are quite well aware of how their attention is distributed. Self-reported data is quite well in line with the objective measures. When a Pearson correlation was computed to assess the relationship between self-reported distribution of attention between the devices and the objective measure made by the eye tracking glasses, a moderate positive correlation was found, $r = .62, n = 74, p < .001$. This is indicating that the participants are well aware what the are paying attention to during a second screen viewing session.

No correlation was found between the number of transition between TV and tablet and the reported multitasking effort. This suggests that the task switch in this kind of multimedia environment is quite inexpensive and effortless.

4.6.1 Tablet Interactions

Undoubtedly, interacting with a device results in an increased cognitive load in comparison with just directing visual attention to it. Two types of participant tablet interactions were recorded: retweets and two types of feed control interactions. A control interaction could either be i) that the feed was stopped by scrolling it downwards or ii) that the feed was scrolled

to the top, allowing new tweets to appear. A retweet was made by clicking on the retweet link which appeared in the tweet.

No clear link between the amount of tablet interaction and eye tracking data was found. Though it is clear that when no content is shown, no interaction can take place and only a negligible amount of visual attention will be directed on the blank second screen device.

Following the findings presented in Section 4.5.3 the participants are categorized into two groups, the ones who chose to retweet and those who did not. Feed control interaction among those who did not retweet was significantly lower ($M = 1.19, SD = 1.75$) in comparison with those who did retweet ($M = 2.97, SD = 3.20$), $t(82) = -3.31, p = .001$. It is worth to note that the feed interaction does not contain the retweets that would increase the difference even further.

Continuing with a similar classification, a statistically significant difference ($t(318) = -2.74, p = .007$) could be seen in the polychronicity score between those who retweeted and those who did not. Those who retweeted had a higher polychronicity score ($M = 41.48, SD = 7.83$) in comparison with those who did not retweet ($M = 38.20, SD = 8.22$). This difference in polychronicity is most likely to explain the retweeting behaviour.

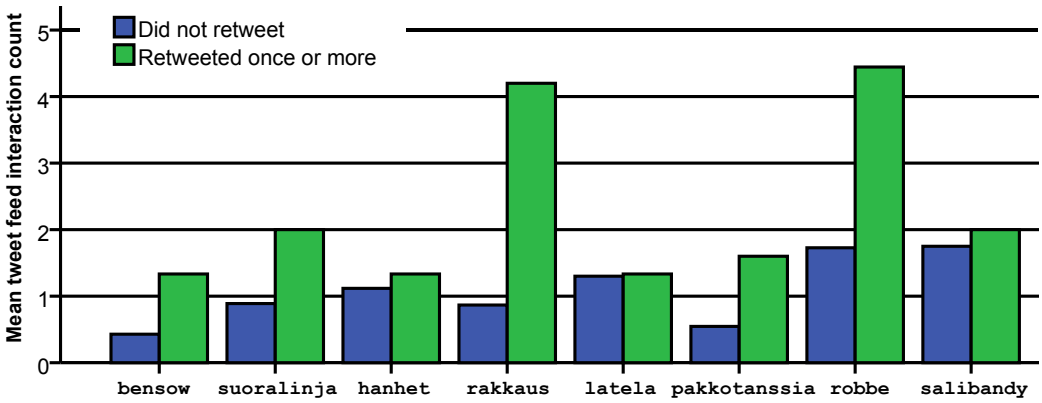


Figure 4.16: Participants who chose to retweet also played around with the feed more.

4.6.2 Distribution of Attention

No significant differences could be found between different TV stimuli regarding the different ETG measures. When comparing the different second screen conditions within TV stimuli, only `robbe` ($F(2, 25) = 6.43, p = .006$) and `salibandy` ($F(2, 25) = 4.70, p = .019$) showed statistically significant

differences between the different situations. The addinfo condition (II) differed again from the situation where both addinfo and tweets were shown (IV). The distribution of attention during different second screen condition is show in Figure 4.17.

Visually some characteristics can be seen between the different genres, but again no statistically significant data was found to support these findings. Nevertheless, magazine, reality, and sport tweets (condition III) seem to gather slightly more interest in comparison with addinfo (condition II). Within the documentary genre, addinfo seems to gather marginally more tablet interest. This can be supported by the fact that documentary tweets scored least in the majority of the measures as shown in Figure 4.14.

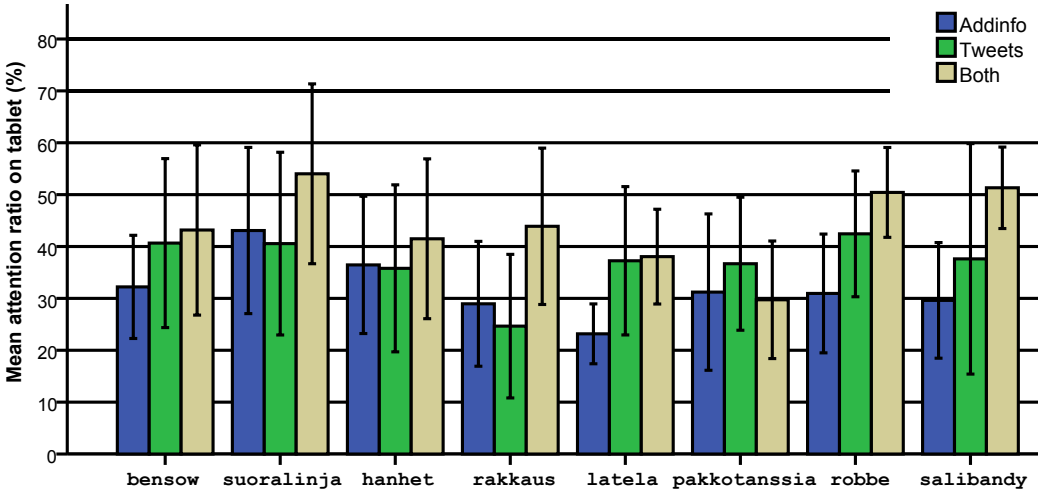


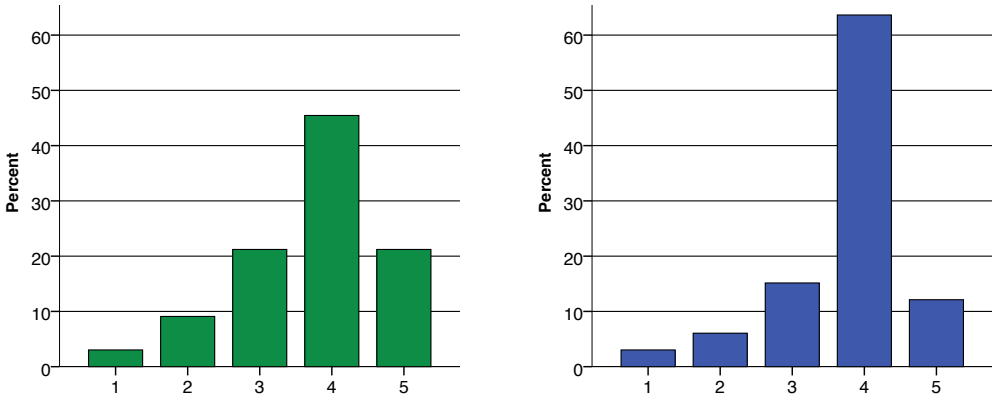
Figure 4.17: Bars represent the proportion of visual attention directed towards the tablet device in the different viewing conditions.

Further, it is noteworthy that second screen condition IV (both tweets and addinfo) scored the highest in all of the stimuli. In some TV conditions even over 50 per cent of the time was spent watching the second screen. It is worth noting that this is the mean score, which implies that some participants spent significantly more time focused on the tablet. Participants' polychronicity score showed a significant yet a weak correlation $(r = .22, n = 204, p = .002)$ between the time spent focused on the tablet, i.e. the focus on the table increased as with higher polychronicity scores. However, no significant correlation was found between the number of transitions made between the two devices and the polychronicity score $(r = .11, n = 204, p = .002)$. A decrease in reported concentration can also be seen when more tablet content is introduced, as pointed out earlier in Section 4.4.

4.7 Companion Application Experience

Once the experimental phase was finished, each participant was asked to more freely describe and comment on the experiment. In addition to Likert-style questions text areas were provided to enable free commenting on things they felt like commenting (Appendix D). In general, the application design and user interface was praised as good. This post experiment questionnaire managed to capture slightly different aspects in comparison to the questionnaires presented after each trial as presented next.

In the open questions participants mentioned that the compaction application was simple to use and that the content was clearly displayed. These findings are supported by highly scored reported tablet pleasantness and tablet usefulness (Figure 4.18).



(a) Tablet pleasantness (b) Tablet usefulness

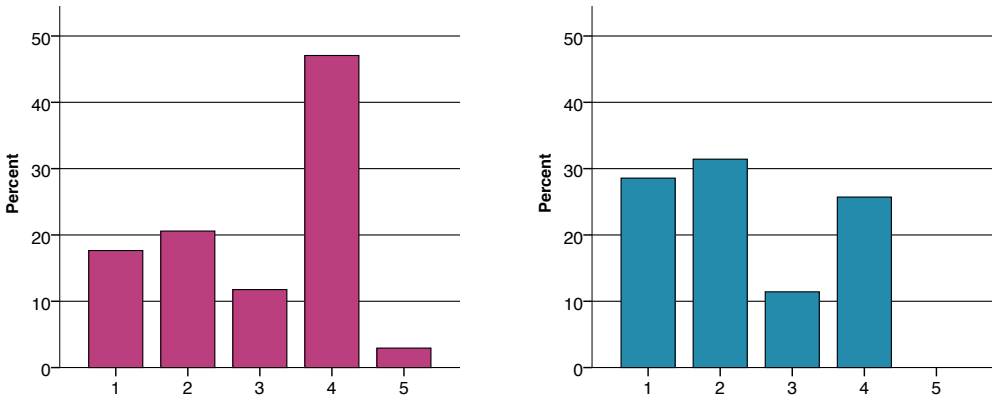
Figure 4.18: After that the trials had finished, the participants evaluated the companion application experience quite high.

The majority of participants (35 of 43) commented that the second screen application and the content provided were for the most part successfully implemented. The automatic information content delivery received mostly positive feedback. But still nearly half (19 of 43) of the participants reported things to improve. In contrast, the participants were not that pleased with the tweet stream, 19 out of the 43 participants had something negative to comment about the tweets. Viewers would have wanted to be able to choose between receiving tweets or not receiving them. When the tweets were visible, the participants would have wanted to choose which keywords were to be included in the tweet stream to avoid off topic content.

Nearly half, 19 out of 43, suggested feature improvements that dealt mainly with flow control and increase in interaction. For example users

would have wanted to be able to do free searches ("do googling") and browsing in the internet. Suggestions also included things such as the ability to control the content filtering and choosing on which basis and what type of second screen content is presented, were also frequently requested. Some suggested a more customizable interface, changing the window and font sizes to match individual preferences.

Further, some participants wanted to be able to close or hide some of the UI components in order to better concentrate on the TV content. This can well be seen in Figure 4.19a. This figure visualizes participant answers when asked whether if they felt that the tablet disturbed them. Out of interest the opposite question was asked, whether if participants felt that TV disturbed them during use of the tablet. This answers are visualized in Figure 4.19b. Main reasons for the disturbance were the highlighting of new items and the fact that the tablet was mounted in the field of view not allowing to viewer to set the device aside as pleased. Two users also felt that customizable window and font sizes would have made it easier to follow both media.



(a) Tablet disturbed TV (b) TV disturbed Tablet

Figure 4.19: The disturbingness of the two screens received mixed opinions.

When finished reading an article, or when the tweet feed was felt overwhelming, it would have been nice to hide the stream or alternatively disable the highlighting of new messages. For the tweets, users would have wanted to choose whether to receive the feed or not, a feature to filter or rank items by popularity, TV channel or by friends. It was also mentioned that it would have nice to lay the tablet aside during an interesting program or at least disable the highlights indicating new tweets of additional information.

Similarly, a participant suggested that vibration or sound feedback could be desirable if some important stuff would appear, e.g. at home viewing conditions where the tablet might have been put aside.

Almost a half of the participants (44%) would have wanted to be able to more actively take part in the social activities, write their own tweets and choose which social media platform to use. Even though TV content providers tightly bundle programs and Twitter together, our participants often verbally wondered while collecting their belongings why there were no other social media (mainly Facebook) provided in the second screen application.

Supporting the concept of *second* screen, participants frequently mention that if the TV content was less interesting the tablet application added more value to the experience compared with situations where TV content alone was interesting enough.

One participant manages to summarize the findings quite well by saying: "In general, the system was well suited for an experimental situation, but in real life it would need a possibility for much more extensive customization.".

Chapter 5

Discussion

This chapter begins by summarizing the most essential findings of this thesis in relation to the research questions presented in Chapter 1. Then, based on the results, recommendations are given for social second screen applications. This is followed by an assessment of the reliability of the results and by a presentation of possible future research topics. Lastly, the implications are discussed.

5.1 Social Presence During Media Viewing

The main research question of this thesis is: *What are the effects of social presence on the viewing experience?*This question was broken down to two subquestions: *How does the social second screen content affect the viewing experience?*and *How do people appraise the new environment?*These questions take a more subjective point of view than previous, research which seems to have focused mainly on studying performance.

Upon tweet introduction, no statistically significant differences in the scoring of the TV content could be seen. No notable differences were found in the scoring of the tweets with respect to, if shown alone or together with additional information. This indicates that the participants were able to clearly distinguish between the different media. The fact that only marginal differences within the media were found in the different conditions did not mean that there were no significant effect in the overall viewing experience.

Upon tablet introduction, the attention shifted notably from the main screen to the second screen across all programs regardless of content interestingness or other factors. This was supported by both objective metrics and participant self-reporting. In the documentary and reality genres the presence of tweets seemed to lower the reported arousal, while in mag-

azine and sports only a very marginal increase could be seen in the reported arousal levels. In addition, introduction of the tablet undoubtedly increased the multitasking load slightly. With the support of these findings it seems evident that second screen viewing has a notable effect on traditional passive viewing.

Answering the second subquestion, especially the post-experiment questionnaire had an important role. The free-form feedback seemed to capture different characteristics regarding the whole second screen experience in comparison with the individual post-trial questionnaires which mainly seemed to capture feedback on the content. The overall second screen experience received positive feedback, while the feedback given during individual trials focused on the shortcomings of the tweet stimulus that was presented. A majority of the participants also found the tablet useful even with a quite simple companion application implemented. These two findings answer the second subquestion. People prefer a second screen viewing environment over traditional passive environments.

In answer to the main research question, second screen viewing gives easy access to multiple supportive media. It enhances the experience and decreases boredom in situations where the main media satiates the viewer. With these findings it can be concluded that the presence of the second screen improves the viewing experience by providing the new breed of media multitaskers what they want.

Differences with previous research were also found. Contrary to the findings of Poposki and Oswald [56], polychronicity could only seldom explain multitasking preferences of the participants. Neither did the BIS and BAS indexes, presented by Carver and White [12]. This might be explained by the fact that media multitasking differs quite much from traditional multitasking situations. Additionally, multitasking, as the term suggests, usually involves a task or a clear goal to be reached. In this experiment the participants were instructed to just enjoy the contents provided as they pleased.

As suggested by Chen [15], U&G seems to be a good way to explain media use. In a free media multitasking environment, people are likely to consume the media which best suits their needs at a given moment. The high penetration of smart phones and a wide variety of applications available in the software markets already gives the consumer plenty of second screen content to choose from. A large number of social applications are already in active use and are well-established among the viewers. This is why the benefits of only relying on traditional social media (tweets) as the backbone for a program specific companion application needs to be questioned. A successful companion application should have other means of

social interaction, as the Twitter application is already there and in use. It is a good thing to include but it should not be the only thing to rely on.

5.2 Recommendations

In regard to the overall findings, paying the utmost attention to the second screen content gathering and filtration is recommended. Unrelated or out-of-sync content in easily noted and condemned by the viewers. An optimal feed speed is crucial is order not to make the viewer uncomfortable with overwhelming content feeds.

Even though the automated content feeds performed satisfactorily there were situations where customization possibilities would have been desirable. In situations were the content feeds did not please, participants asked about the possibility to be able to control the content feed flow. Such things are the speed of the feed and relevance of the content. Feed speed is set by the amount of content being displayed. Content relevance is determined by the keywords or other means the content is searched for and gathered. It would be preferable to provide such means of customization when introducing companion applications like this. Such small irritating details in the overall experience may ruin even a otherwise good companion application.

Additionally, this study brought forth the social media preferences of Finnish media users (Sections 2.4.2 and 4.5.1). Regardless of its popularity abroad, in Finland Twitter has not seen the same level of popularity as Facebook, which is the number one choice of social media according to our participants. When designing applications like this, paying attention on how to best bring along the favourite platforms to the second screen is recommended. Now when Facebook offers hashtags [32], this may invoke social involvement in countries with low Twitter user penetration. It is noteworthy that hashtag support was introduced after agreeing on the experiment design of this thesis. In addition, Facebook privacy policies still prevent global public access to the user posts, in contrast to Twitter.

Additionally, during a TV show there is seldom time to commit to the tasks requiring increased cognitive load, such as tweeting. These heavy tasks are usually done during ad breaks or post-broadcast. Increasing commitment and providing social content should be done by exploiting program specific characteristics and providing simple means of interactions, e.g. rating, voting, "liking", shearing, etc. And of course there are plenty of possibilities to expand to during pre-release and post-broadcast, in other words, between episodes. Even if not second screen viewing, this would

keep the viewer committed and possibly encourage the use of the second screen application during the broadcast.

As stated before, the media scene and the technical solutions evolve constantly. To ensure relevant and useful second screen content, market research should be continuously performed to ensure that the application is able to fulfill the needs of the modern media multitasker.

5.3 Evaluation and Validity

As with all scientific studies, also here limitations and validity must be discussed. This is to understand how reliable and applicable the results are. Limitations of this work can be approached by studying the validity of this thesis. This can be a split into internal and external validity. External validity refers to the limitations of a generalization of the results. Internal validity refers to the outcome and interpretation of the results. Furthermore, the limitations of the study should be well known and understood, as they may have an effect on how generalizable the results are.

First and foremost, the reliability of a study has to be established to ensure the reproducibility of the results obtained. This is in principle established through the use of relevant stimulus and reliable data and measurements. We can state that statistically significant differences were found between the different program genres. In addition, statistics confirmed that the two programs within each genre are consistent with each other. The four genres could further be grouped into to two types of program with notable differences: fact and entertainment. Thus the stimulus can be considered as reliable. The with-in genre confidence margins could perchance been smaller if the stimuli samples would have been extracted from the same show. Based on the extensive literature survey, the author argues that appropriate tools were used, which in turn assures that the data recorded is reliable.

External validity might be questioned by the participant demographics. The plan was to use rather young participants who are already used to multitasking and who are fast learners. In addition, our participants were largely technically oriented. This might raise concerns that our sample group did not truthfully represent the typical TV consuming population. In contrast, these users are the ones most likely to first adapt this kind of new application.

Internal validity might be questioned by referring to the number of participants. In the light of a statistical power analysis, the number of participants (43) might seem a bit low. In addition to being in line with a typical

participant group size seen in previous studies [33] and recommendations [57], the methods used were able to recognize statistically significant differences between the experimental conditions.

5.4 Future Research

This thesis was limited to a single setup with limitations on how many research questions can be addressed during a single trial while simultaneously keeping the number of participants within manageable limits. This always forces to narrow down the scope of the research. The results and findings of this study raised the following possible future research questions:

I *How to automate the gathering and filtration of social content to provide the best viewing experience possible?*

II *What is the optimal feed speed and how the feed speed is related to the genre in question?*

III *Could other social media outperform the industry standard Twitter?*

IV *How would increased involvement affect the experience?*

In addition, the vivid technological development and rapid changes in the media scene causes continual emergence of new interesting topics for future research. As Rideout et al. [60] stated, passive TV viewing seems to be a thing from the past. Even more radical changes can be seen ahead, and the future of traditional broadcast TV can even be questioned. As Edwards [21] stated in a recent news article, the concept of subscription TV is dying. The amount of cable TV subscribers is declining rapidly, as people move away from watching television towards watching videos. This can also be seen in the rise of online subscription video services such as Netflix[1]. This sets its own challenges for second screen companion applications as well. It remains to be seen when a transition will take place causing the second screen to become the primary one.

[1] `http://www.netflix.com`

Chapter 6

Conclusions

The objective of this thesis was to expand knowledge of second screen media experiences. Existing literature have left room for affective research, as most of the previous publications focus on performance. The thesis sought differences between single task (TV only) and multitasking viewing conditions. In the multitasking condition viewers were provided with social content on a secondary screen to accompany the TV content. To evaluate the differences, both subjective and objective measurements were used.

The following three research questions were presented: i) *What are the effects of social presence on the viewing experience?* i.i) *How does the social second screen content affect the viewing experience?* and i.ii) *How do people appraise the new environment?* All these were successfully answered based on the findings of an experimental study. In order to answer the questions, a thorough literature review was accomplished with regard to different media environments and the types of studies to be performed, ranging from methods to practices.

Based on the results and the discussion, the conclusion is that even though the tweets did not affect how the TV content was appraised, the presence of the second screen improved the participants' viewing experience. Moreover, it seems that entertainment programs benefit from the social second screen content more than fact programs do. Both relevance of the tweets and the speed at which the content was presented had an effect on how well the second screen content was received by the viewers. In situations where the second screen did not score that well, the content was often said to be out-of-sync or irrelevant for the viewer or with the content in mind. This was a content related issue, not a direct failure of the second screen.

To conclude, this study brought forth the effects of Twitter based social presence in an environment where Twitter has not reached the same level

of penetration as it has in other countries where second screen studies have been made. Nevertheless, the results are encouraging. In comparison with studies focusing only on human multitasking performance, positive effects of second screen viewing could be distinguished and recommendations based on these findings suggested. TV content is here to stay, and so is media multitasking and social networking. Yet, it remains to be seen how and through which physical media the contents are delivered and presented to the viewer.

Bibliography

[1] Rachel F. Adler and Raquel Benbunan-Fich. Juggling on a high wire: Multitasking effects on performance. *International Journal of Human-Computer Studies*, 70(2):156–168, 2012.

[2] John J.B. Allen. Calculating metrics of cardiac chronotropy: A pragmatic overview. *Psychophysiology*, 39:S18, 2002.

[3] Erik M. Altmann and Gregory J. Trafton. Memory for goals: An activation-based model. *Cognitive science*, 26(1):39–83, 2002.

[4] Rafael Ballagas. "Media Multitasking" is not always Multitasking. *Nokia Research Center*, June 2009.

[5] Gary G. Berntson, Thomas J. Bigger, Dawin L. Eckberg, Paul Grossman, Peter G. Kaufmann, Marek Malik, Haikady N. Nagaraja, Stephen W. Porges, Philip J. Saul, Peter H. Stone, and Maurits W. van der Molen. Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology*, 34(6):623–648, November 1997. ISSN 0048-5772.

[6] Allen C. Bluedorn and Kimberly S. Jaussi. Organizationally relevant dimensions of time across levels of analysis. *Research in Multi Level Issues*, 6:187–223, 2007.

[7] Margaret M. Bradley and Peter J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[8] Adam S. Brasel and James Gips. Media multitasking behavior: Concurrent television and computer usage. *Cyberpsychology, Behavior, and Social Networking*, 14(9):527–534, 2011.

[9] Jennings Bryant and Dolf Zillmann. Using television to alleviate boredom and stress: Selective exposure as a function of induced excitational states. *Journal of Broadcasting & Electronic Media*, 28(1):1–20, 1984.

[10] Ann-Frances Cameron and Jane Webster. Relational outcomes of multicommunicating: Integrating incivility and social exchange perspectives. *Organization Science*, 22(3):754–771, 2011.

[11] Mark L. Carrier, Nancy A. Cheever, Larry D. Rosen, Sandra Benitez, and Jennifer Chang. Multitasking across generations: Multitasking choices and difficulty ratings in three generations of americans. *Computers in Human Behavior*, 25(2):483–489, 2009.

[12] Charles S. Carver and Teri L. White. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The bis/bas scales. *Journal of Personality and Social Psychology*, 67(2):319, 1994.

[13] Pablo Cesar, Dick C. A. Bulterman, and A.J. Jansen. Usages of the secondary screen in an interactive television environment: control, enrich, share, and transfer television content. In *Changing television environments*, pages 168–177. Springer, 2008.

[14] Pablo Cesar, Konstantinos Chorianopoulos, and Jens F. Jensen. Social television and user interaction. *Computers in Entertainment (CIE)*, 6(1): 4, 2008.

[15] Gina Masullo Chen. Tweet this: A uses and gratifications perspective on how active twitter use gratifies a need to connect with others. *Computers in Human Behavior*, 27(2):755–762, 2011.

[16] Sangwon Choi, Jaehyun Han, Geehyuk Lee, Narae Lee, and Woohun Lee. Remotetouch: touch-screen-like interaction in the tv viewing environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 393–402. ACM, 2011.

[17] James T. Croasmun and Lee Ostrom. Using likert-type scales in the social sciences. *Journal of Adult Education*, 40(1):19–22, 2011.

[18] Leon Cruickshank, Emmanuel Tsekleves, Roger Whitham, Annette Hill, and Kaoruko Kondo. Making interactive tv easier to use: Interface design for a second screen approach. *The Design Journal*, 10(3): 41–53, November 2007.

[19] Richard L. Daft and Robert H. Lengel. Organizational information requirements, media richness and structural design. *Management Science*, 32(5):554–571, 1986.

[20] Evelien D'heer, Cédric Courtois, and Steve Paulussen. Everyday life in (front of) the screen: the consumption of multiple screen technologies in the living room context. In *Proceedings of the 10th European conference on Interactive tv and video*, pages 195–198. ACM, 2012.

[21] Jim Edwards. Tv is dying, and here are the stats that prove it. Webpage — businessinsider.com, November 2013. URL `http://www.businessinsider.com/cord-cutters-and-the-death-of-tv-2013-11`. Accessed 24.11.2013.

[22] Encyclopædia Britannica Online. Encyclopædia Britannica Online Academic Edition — Human Nervous System. Webpage, August 2013. URL `http://www.britannica.com/EBchecked/topic/409709/human-nervous-system`. Accessed 19.8.2013.

[23] Evolumedia group. The second screen and television - Overview and growth perspectives. White paper series No 1. Canada Media Fund, October 2012.

[24] Evolumedia group. The second screen and television - Benefits & impacts. White paper series No 2. Canada Media Fund, March 2013.

[25] Erving Goffman. The presentation of self in everyday life. *New York: Anchor Doubleday*, 212:1–17, 1959.

[26] Jan Hess, Benedikt Ley, Corinna Ogonowski, Lin Wan, and Volker Wulf. Jumping between devices and services: towards an integrated concept for social tv. In *Proceddings of the 9th international interactive conference on Interactive television*, pages 11–20. ACM, 2011.

[27] International Telecommunication Union - Radiocommunication Sector. General viewing conditions for subjective assessment of quality of sdtv and hdtv television pictures on flat panel displays. *Recommendation ITU-R BT.2022*, August 2012.

[28] Andreas M. Kaplan and Michael Haenlein. Users of the world, unite! the challenges and opportunities of social media. *Business horizons*, 53 (1):59–68, 2010.

[29] Elihu Katz, Jay G. Blumler, and Michael Gurevitch. Uses and gratifications research. *The Public Opinion Quarterly*, 37(4):509–523, 1973.

[30] Cornelius J. König and Mary J. Waller. Time for reflection: A critical examination of polychronicity. *Human Performance*, 23(2):173–190, 2010.

[31] Kenta Kusumoto. Factors in additional information affecting media experience of tv second screen application. Master's thesis, Aalto University School of Science, Espoo, Finland, January 2014.

[32] Justin Lafferty. A closer look at facebook's hashtags. Webpage — Insidefacebook.com, November 2013. URL `http://www.insidefacebook.com/2013/06/17/a-closer-look-at-facebooks-hashtags/`. Accessed 20.11.2013.

[33] Mari Laine-Hernandez, Teemu Kinnunen, Kenta Kusumoto, Jari Kätsyri, Stina Westman, Niklas Ravaja, and Pirkko Oittinen. Multitasking influence on engagement. In *Personal Media Day, Deliverable 2.3.3.7*, Nextmedia, pages 1–35, April 2013.

[34] Annie Lang. The limited capacity model of mediated message processing. *Journal of Communication*, 50(1):46–70, 2000.

[35] Robert LaRose, Dana Mastro, and Matthew S. Eastin. Understanding internet usage a social-cognitive approach to uses and gratifications. *Social Science Computer Review*, 19(4):395–413, 2001.

[36] Jeff T Larsen, Catherine J Norris, and John T Cacioppo. Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, 40(5):776–785, 2003.

[37] Amanda Lenhart, Kristen Purcell, Aaron Smith, and Kathryn Zickuhr. Social media and young adults. *Pew Internet & American Life Project*, 3, 2010.

[38] Robert H. Logie, Steven Trawley, and Anna Law. Multitasking: multiple, domain-specific cognitive functions in a virtual environment. *Memory & cognition*, 39(8):1561–1574, 2011.

[39] Nora Madjar and Christina E. Shalley. Multiple tasks' and multiple goals' effect on creativity: Forced incubation or just a distraction? *Journal of Management*, 34(4):786–805, 2008.

[40] Glynn W. Mangold and David J. Faulds. Social media: The new hybrid element of the promotion mix. *Business Horizons*, 52(4):357–365, 2009.

[41] Alice E. Marwick et al. I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1):114–133, 2011.

[42] Daniel McDonald, Jingbo Meng, and Sharon Kleinman. The multitasking of entertainment. *The Culture of Efficiency: Technology in Everyday Life. New York: Peter Lang*, pages 142–157, 2009.

[43] Hugh Miller. The presentation of self in electronic life: Goffman on the internet. In *Embodied Knowledge and Virtual Space Conference*, volume 9, 1995.

[44] Kathryn Montgomery. Youth and digital media: A policy research agenda. *Journal of Adolescent Health*, 27(2):61–68, 2000.

[45] Carlos H. Morimoto and Marcio R. M. Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1):4–24, 2005.

[46] Hendrik Müller, Jennifer Gove, and John Webb. Understanding tablet use: a multi-method exploration. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 1–10. ACM, 2012.

[47] Jakob Nielsen. *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993. ISBN 0-12-518405-0.

[48] Charles Robert Noback, Norman L. Strominger, Robert John Demarest, and David A. Ruggiero. *The Human Nervous System: Structure and Function*. Number 744 in SpringerLink: Springer e-Books. Humana Press, 2005.

[49] Domen Novak, Matjaž Mihelj, and Marko Munih. Dual-task performance in multimodal human-computer interaction: a psychophysiological perspective. *Multimedia Tools and Applications*, 56(3):553–567, 2012.

[50] Official Statistics of Finland. Social media: use of online media and social network services. Webpage, October 2010. URL `http://www.stat.fi/til/sutivi/2010/sutivi_2010_2010-10-26_kat_003_fi.html`. Accessed 4.10.2013.

[51] Official Statistics of Finland. Developments in internet use. Webpage, November 2012. URL `http://tilastokeskus.fi/til/sutivi/2012/sutivi_2012_2012-11-07_kat_001_fi.html`. Accessed 4.10.2013.

[52] Official Statistics of Finland. Use of social networking services. Webpage, November 2012. URL `http://tilastokeskus.fi/til/sutivi/2012/sutivi_2012_2012-11-07_kat_004_fi.html`. Accessed 4.10.2013.

[53] Lucy Jo Palladino. *Find Your Focus Zone: An Effective New Plan to Defeat Distraction and Overload*. simonandschuster.com, 2007.

[54] Stephen J Payne, Geoffrey B Duggan, and Hansjörg Neth. Discretionary task interleaving: heuristics for time allocation in cognitive foraging. *Journal of Experimental Psychology: General*, 136(3):370, 2007.

[55] Don E Pearson. Viewer response to time-varying video quality. In *Photonics West'98 Electronic Imaging*, pages 16–25. International Society for Optics and Photonics, 1998.

[56] Elizabeth M. Poposki and Frederick L. Oswald. The multitasking preference inventory: Toward an improved measure of individual differences in polychronicity. *Human Performance*, 23(3):247–264, 2010.

[57] Niklas Ravaja. Contributions of psychophysiology to media research: Review and recommendations. *Media Psychology*, 6(2):193–235, 2004.

[58] Lamar N. Reinsch, Jeanine Warisse Turner, and Catherine H. Tinsley. Multicommunicating: A practice whose time has come? *Academy of Management Review*, 33(2):391–403, 2008.

[59] Jennifer C. Richardson and Karen Swan. Examining social presence in online courses in relation to students' perceived learning and satisfaction. *Journal of Asynchronous Learning Networks*, 7(1):68–88, 2003.

[60] Victoria J. Rideout, Ulla G. Foehr, and Donald F. Roberts. Generation M^2: Media in the lives of 8- to 18-year-olds, 8 2010.

[61] Joshua S. Rubinstein, David E. Meyer, and Jeffrey E. Evans. Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, 27(4):763, 2001.

[62] Thomas E. Ruggiero. Uses and gratifications theory in the 21st century. *Mass communication & society*, 3(1):3–37, 2000.

[63] Aaron M Sackett, Tom Meyvis, Leif D Nelson, Benjamin A Converse, and Anna L Sackett. You're having fun when time flies the hedonic consequences of subjective time progression. *Psychological Science*, 21 (1):111–117, 2010.

[64] Dario D. Salvucci and Joseph H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, pages 71–78. ACM, 2000.

[65] Dario D. Salvucci and Niels A. Taatgen. *The multitasking mind*. Oxford University Press, 2010.

[66] Samsung Electronics Co., Ltd. Galaxy tab 10.1 - samsung mobile. Webpage, July 2013. URL `http://www.samsung.com/global/microsite/galaxytab/10.1/spec.html`. Accessed 26.7.2013.

[67] Hope Jensen Schau and Mary C. Gilly. We are what we post? self-presentation in personal web space. *Journal of Consumer Research*, 30 (3):385–404, 2003.

[68] Stefan Schmidt and Harald Walach. Electrodermal activity (eda)-state-of-the-art measurement and techniques for parapsychological purposes. *Journal of Parapsychology*, 64(2): 139–164, 2000. URL `http://www.biomedsearch.com/article/Electrodermal-activity-eda-state-art/67718670.html`.

[69] SensoMotoric Instruments (SMI). Smi eye tracking glasses - discover what is seen. Webpage, July 2013. URL `http://www.eyetracking-glasses.com/fileadmin/user_upload/documents/smi_etg_flyer.pdf`. Accessed 29.7.2013.

[70] John Short, Ederyn Williams, and Bruce Christie. The social psychology of telecommunications. 1976.

[71] Heloisa Simon, Eros Comunello, and Aldo von Wangenheim. Enrichment of interactive digital tv using second screen. *International Journal of Computer Applications*, 64(22):58–64, 2013.

[72] Sonera Oyj. Soneran koti ja tv –tutkimus 2012. Webpage, 2013. URL `http://mb.cision.com/Public/145/9298248/b709e1fc921675e2.pdf`. Accessed 5.11.2013.

[73] Jatin Srivastava. Media multitasking performance: Role of message relevance and formatting cues in online environments. *Computers in Human Behavior*, 29(3):888–895, 2013.

[74] Statistics of Finland. Changes in use of time. Webpage, 2009. URL `http://www.stat.fi/til/akay/2009/02/akay_2009_02_2011-02-17_kat_002_fi.html`. Accessed 4.10.2013.

[75] The Nielsen Company. New study confirms correlation between twitter and tv ratings. Webpage, March 2013. URL `http://www.nielsen.com/us/en/newswire/2013/new-study-confirms-correlation-between-twitter-and-tv-ratings.html`. Accessed 23.4.2013.

[76] Mike Thelwall and Kevan Buckley. Topic-based sentiment analysis for the social web: The role of mood and issue-related words. *Journal of the American Society for Information Science and Technology*, 2013.

[77] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, 2010.

[78] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. Sentiment in twitter events. *Journal of the American Society for Information Science and Technology*, 62(2):406–418, 2011.

[79] Fatih Tokan, Pekka Mattila, and Elina Koivisto. Media as multitasking: An exploratory study on capturing audiences' media multitasking and multiple media use behaviours. *Project report. Aalto University School of Economics.*, Oct 2011. URL `http://www.aaltomediamark.org/wp-content/uploads/2011/04/Media-as-Multitasking-NextMedia.pdf`. Accessed 5.11.2013.

[80] TwitterUK. Tune In With Twitter — Driving Discovery and Engagement with TV. URL `https://tweet.twitter.com/TVbook`, 2013 January.

[81] Anton van Boxtel. Facial emg as a tool for inferring affective states. In *Proceedings of Measuring Behavior*, pages 104–108, 2010.

[82] Maria Viitanen, Stina Westman, Teemu Kinnunen, and Pirkko Oittinen. Features of media multitasking experiences nextmedia deliverable 2.3.3.6. In *Media Multitasking*, Nextmedia, pages 1–41, August 2012.

[83] Zheng Wang and John M. Tchernev. The "myth" of media multitasking: Reciprocal dynamics of media multitasking, personal needs, and gratifications. *Journal of Communication*, 62(3):493–513, 2012.

[84] Jeffrey Wilson. 5 great second-screen apps to enhance the tv experience. Webpage, October 2013. URL `http://www.pcmag.com/article2/0,2817,2425396,00.asp`. Accessed 13.1.2014.

[85] Jinxi Xu and Bruce W. Croft. Query expansion using local and global document analysis. In *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 4–11. ACM, 1996.

[86] Robert M. Yerkes and John D. Dodson. The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5):459–482, 1908.

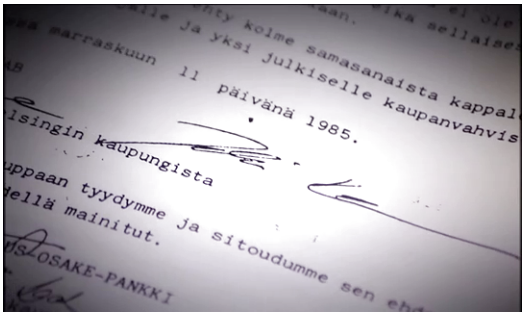
[87] YLE, Finnish Broadcasting Company. Twitter-teksti-tv - mikä se on? Webpage, February 2012. URL `http://yle.fi/tekstitv/arkisto/mika_teksti-tv/twitter-teksti-tv_-_mika_se_on_1454.html`. Accessed 23.4.2013.

[88] YLE, Finnish Broadcasting Company. Suoran linjan nettikeskustelun pelisäännöt. Webpage, April 2012. URL `http://yle.fi/uutiset/suoran_linjan_nettikeskustelun_pelisaannot/5881832`. Accessed 23.4.2013.

[89] Weiyu Zhang and Lingzi Zhang. Explicating multitasking with computers: Gratifications and situations. *Computers in Human Behavior*, 28 (5):1883–1891, 2012.

Appendix A

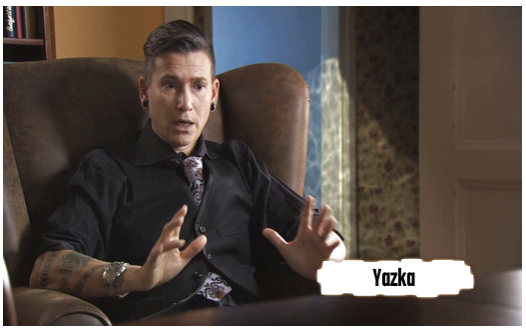
Video Stimulus



MOT: Bensowin säätiön salaisuus
Aired 15.4.2013 at 20:00
Duration: 4:55 — 34 tweets
Magazine. Exposes the financial abuse done by Bensow childcare foundation board members.

YLE Suora linja
Aired 9.4.2013 at 19:20
Duration: 5:03 — 53 tweets
Magazine. Interactive news program discussing the mistakes done by mining company Talvivaara and the contradictory sacking of elementary school teacher.

Avara luonto: Stadin valkoposket
Aired 27.4.2013 at 18:45
Duration: 4:37 — 35 tweets
Documentary. Tell the story of how the Barnacle Goose has invaded Helsinki and how people react to that.



Totuus rakkaudesta
Aired 19.4.2013 at 20:00
Duration: 5:25 — 54 tweets
Documentary. Famous people tell their views on everyday situations in a relationship. This episode discusses couple moving together.

Latela
Aired 15.4.2013 at 19:30
Duration: 5:01 — 33 tweets
Reality. Car tuner team from Tampere region compete in fixing and tuning old cars with a tight budget and limited time. This is the final episode where they get to be judged at American Car Show.

Pakko tanssia
Aired 6.4.2013 at 20:00
Duration: 5:00 — 63 tweets
Reality. Dancing competition in search for the best dancing performance in Finland. This sample features the performance of Suma Ensambles and some funny lines from the judges.

Salibandyn EFT:
Finland vs. Sweden
Aired 28.4.2013 at 16:10
Duration: 5:15 — 62 tweets
Sports. Men's Euro Floor ball Tour: Finland vs Sweden. This sample is from the 2nd round. Sweden is leading 0-3. Clip ends by Finland making a goal leaving the situation at 2-4.

Boxing: Robert Helenius vs. Michael Sprott
Aired 23.3.2013 at 22:30
Duration: 5:58 — 63 tweets
Sports. Boxing competition taking place in Germany. The clip covers end of the intro and the complete first round. Finnish commentary.

Appendix B

Background Questionnaire

Perustiedot / *Basic info*

1. Sukupuolesi
   Sex

2. Ikä
   Age

3. Oletko oikea- vai vasenkätinen?
   Are you left or right handed?

4. Jos sinulla on alentunut kuulo- tai näkökynnys, merkitse mitkä.
   Do you have impaired hearing or vision?

5. Koulutus
   Education

6. Koulutus- tai ammattiala
   Field of study or profession

7. Arvioi kuinka paljon käytät tietokonetta keskimäärin päivän aikana?
   Estimate your average daily computer use?

8. Kuinka paljon käytät eri medioita keskimäärin päivän aikana?
   How much do you consume media during the day?

9. Omistatko jonkin tablet-laitteen?
   Jos omistat, arvioi kuinka paljon käytät sitä keskimäärin.
   Do you own a tablet device?
   In case you do, estimate how much you use it during the day.

10. Tyypillisesti TV:tä katsoessani käytän seuraavia laitteita:
    While watching TV, I typically use the following devices:
    ✓ Answered by choosing any of the following:
    TV / Desktop compute / Laptop / Mobilephone / Tablet computer / I don't own a TV / I feel that I don't consume TV media.

11. Kun katson TV:tä äly-laitteen kanssa, koen että päämedia on…
    While simultaneously watching TV accompanied by an other device, I feel that the main media is …
    ✓ Answered by choosing one of the following:
    usually TV. / usually a smart device.

Polikronisuus / *Polycronicity*

Answers are given on the following scale:
täysin eri mieltä *strongly disagree* **1 … 5** täysin samaa mieltä *strongly agree*

12. Työskentelen mieluiten usean projektin parissa päivittäin, sen sijaan että tekisin yhden projektin valmiiksi ja siirtyisin sitten seuraavaan.
    I prefer to work on several projects in a day, rather than completing one project and then switching to another.

13. Haluaisin työn, jossa siirtyisin jatkuvasti tehtävästä toiseen, kuten vastaanottovirkailija tai lennonjohtaja.
    I would like to work in a job where I was constantly shifting from one task to another, like a receptionist or an air traffic controller.

14. Menetän kiinnostukseni tekemiseni kohteeseen, jos joudun keskittymään pitkiä aikoja ilman että ajattelen tai teen jotain muuta.
    I lose interest in what I am doing if I have to focus on the same task for long periods of time, without thinking about or doing something else.

15. Kun teen useita tehtäviä, vaihtelen niiden välillä mieluummin kuin teen yhden kerrallaan.
    When doing a number of assignments, I like to switch back and forth between them rather than do one at a time.

16. Haluan saada yhden tehtävän päätökseen ennen kuin keskityn mihinkään muuhun.
    I like to finish one task completely before focusing on anything else.

17. Oloni on epämukava, jos en voi suorittaa yhtä tehtävää loppuun ennen kuin keskityn toiseen tehtävään.
It makes me uncomfortable when I am not able to finish one task completely before focusing on another task.

18. Olen uppoutuneempi tekemiseeni, jos voin vaihdella usean eri tehtävän välillä.
I am much more engaged in what I am doing if I am able to switch between several different tasks.

19. En pidä siitä, että minun tulee jakaa huomioni usean tehtävän kesken.
I do not like having to shift my attention between multiple tasks.

20. Vaihtelen mieluummin usean projektin välillä kuin keskitän ponnisteluni vain yhteen.
I would rather switch back and forth between several projects than concentrate my efforts on just one.

21. Työskentelisin mieluiten ympäristössä, jossa voin saada valmiiksi yhden tehtävän ennen kuin aloitan seuraavan.
I would prefer to work in an environment where I can finish one task before starting the next.

22. En pidä siitä, kun minun täytyy kesken tehtävän tehdä jotain muuta.
I don't like when I have to stop in the middle of a task to work on something else.

23. Kun minulla on tehtävä suoritettavana, haluan jakaa sen siirtymällä ajoittain muihin tehtäviin.
When I have a task to complete, I like to break it up by switching to other tasks intermittently.

24. Ajattelen vain yhtä asiaa kerrallaan.
I have a "one-track" mind.

25. En halua, että minua keskeytetään kun teen tehtävääni.
I prefer not to be interrupted when working on a task.

BIS / BAS

Answers are given on the following scale:

ei pidä ollenkaan paikkaansa minun kohdallani *very false for me*	**1 … 4**	pitää täysin paikkansa minun kohdallani *very true for me*

26. Vaikka minulle olisi tapahtumassa jotakin ikävää, tunnen itseni vain harvoin pelokkaaksi tai hermostuneeksi.
Even if something bad is about to happen to me, I rarely experience fear or nervousness.

27. Olen valmis näkemään paljon vaivaa saadakseni sen mitä haluan.
I go out of my way to get things I want.

28. Kun jokin asia sujuu minulta hyvin, jatkan erittäin mielelläni sen tekemistä.
When I'm doing well at something I love to keep at it.

29. Olen aina halukas kokeilemaan jotakin uutta, mikäli uskon sen olevan hauskaa.
I'm always willing to try something new if I think it will be fun.

30. Kun saan jotakin mitä haluan, tunnen itseni innostuneeksi ja energiseksi.
When I get something I want, I feel excited and energized.

31. Saamani kritiikki tai moitteet pahoittavat mieltäni aika tavalla.
Criticism or scolding hurts me quite a bit.

32. Halutessani jotakin teen yleensä kaikkeni saadakseni sen.
When I want something I usually go all-out to get it.

33. Teen usein asioita vain sen vuoksi, että ne voivat olla hauskoja.
I will often do things for no other reason than that they might be fun.

34. Jos huomaan tilaisuuden saada jotakin mitä haluan, toimin välittömästi saadakseni sen.
If I see a chance to get something I want I move on it right away.

35. Olen melko huolissani tai järkyttynyt, mikäli luulen tai tiedän jonkun olevan vihainen minulle.
I feel pretty worried or upset when I think or know somebody is angry at me.

36. Kun näen mahdollisuuden johonkin mistä pidän, tulen heti kiihtyneeksi.
When I see an opportunity for something I like I get excited right away.

37. Toimin usein hetken mielijohteesta.
    I often act on the spur of the moment.

38. Jos ajattelen, että jotakin epämielyttävää tulee tapahtumaan, tulen tavallisesti varsin hermostuneeksi.
    If I think something unpleasant is going to happen I usually get pretty "worked up."

39. Kun minulle tapahtuu jotakin hyvää, se vaikuttaa minuun voimakkaasti.
    When good things happen to me, it affects me strongly.

40. Tunnen itseni huolestuneeksi, kun koen suoriutuneeni huonosti jossakin tärkeässä asiassa.
    I feel worried when I think I have done poorly at something important.

41. Janoan jännitystä ja uusia kokemuksia.
    I crave excitement and new sensations.

42. Kun tavoittelen jotakin, niin mikään ei pidättele minua.
    When I go after something I use a "no holds barred" approach.

43. Pelkään hyvin harvoja asioita verrattuna ystäviini.
    I have very few fears compared to my friends.

44. Minusta olisi jännittävää voittaa jokin kilpailu.
    It would excite me to win a contest.

45. Olen huolissani siitä, että saatan tehdä virheitä.
    I worry about making mistakes.

Sosiaalinen media / *Social Media*

Answers are given on the following scale:
(unless stated otherwise)

täysin eri mieltä *strongly disagree* **1 … 5** täysin samaa mieltä *strongly agree*

46. Koen käyttäväni aktiivisesti seuraavia sosiaalisia medioita:
I actively use the following social media:
✓ Aswered by choosing any of the following:
Facebook, Google+, Twitter, Formspring, Habbo, LinkedIn, Other

47. Mitä sosiaalinen läsnäolo merkitsee sinulle?
What does social presence signify for me?
✎ Answered in writing.

48. Sosiaalinen media on minulle…
Social media means the following to me…
✎ Answered in writing.

49. Sosiaalisen median kautta ylläpidän suhteita seuraaviin ryhmiin:
Through social media, I maintain relationships with the following groups:
✓ Answered by choosing any of the following:
firends / romantic realtions / family relation /
social cirlces / community relations / religious organizations

50. Kun kulutan sosiaalista mediaa, koen tärkeimmäksi…
Using social media, I feel the most important thing is …
✓ Answered by choosing any of the following:
follow what others are doing / share my own experiences /
comments on others updates / do private messaging.

51. Kun kulutan sosiaalista mediaa, koen toiseksi tärkeimmäksi…
Using social media, I feel the second most important thing is …
✓ Answered by choosing any of the following:
follow what others are doing / share my own experiences /
comments on others updates / do private messaging.

52. Kun kulutan sosiaalista mediaa, koen kolmanneksi tärkeimmäksi…
Using social media, I feel the third most important thing is …
✓ Answered by choosing any of the following:
follow what others are doing / share my own experiences /
comments on others updates / do private messaging.

53. Tunnen olevani osa Twitter-yhteisöä.
    I feel like being part of the Twitter community.

54. Muita kiinnostaa, mitä jaan sosiaalisessa mediassa.
    A care about what I share on social media.

55. Sosiaalisessa mediassa luon sisältöä kavereitani ajatellen.
    In social media I create content with my friends in mind.

56. Sosiaalisessa mediassa luon sisältöä tuntemattomille.
    In social media, I create content with strangers in mind.

Appendix C

Post-Trial Questionnaire

Affective meters

Answers are given using the 9 step SAM scale (see Bradley and Lang [7]).

1. Arvioi mielipaha vs. mielihyvä - millaisen tunteen äsken nähty TV-ohjelma ja tabletilla esitetty sisältö herättivät sinussa?
   Evaluate the TV and tablet experience on a scale rangeing from pleased to unsatisfied. What where your feelings during the viewing session?

2. Arvioi emotionaalinen aktivaatio - millaisen tunteen äsken nähty TV-ohjelma ja tabletilla esitetty sisältö herättivät sinussa?
   Evaluate your emotional activation during the viewing session.

TV-ohjelma / *TV-program*

Answers are given on the following scale:

täysin eri mieltä *strongly disagree* **1 … 7** täysin samaa mieltä *strongly agree*

3. TV-ohjelma oli mielenkiintoinen.
   TV program was interesting

4. Pidin TV-ohjelmasta.
   I liked the TV program.

5. Keskityin TV-ohjelmaan.
   I focused on the TV program.

6. Uppouduin TV-ohjelmaan täydellisesti.
   I immersed myself into the TV program.

7. Jakaisin TV-ohjelman ystävilleni.
   I would share the program with my friends.

8. Voisin ostaa tämän yksittäisen TV-ohjelman.
   I could consider buying this TV show.

Twiitit / *Tweets*

Answers are given on the following scale:
täysin eri mieltä *strongly disagree* **1 … 7** täysin samaa mieltä *strongly agree*

9. Twiitit olivat mielenkiintoisia.
   The tweets were interesting.

10. Pidin twiiteistä.
    I liked the tweets.

11. Uppouduin twiitteihin täydellisesti.
    I immersed myself into the tweets.

12. Tweettien lukeminen toi ohjelmaan lisäarvoa.
    Reading the tweets brought added value to the show.

13. Katselukokemus oli mielyttävämpi kun luin tweettejä.
    Viewing experience was more pleasant while reading the tweets.

14. Tweetit liittyivät ohjelmaan.
    The tweets were related to the show.

15. Tweetit saivat minut tuntemaan itseni sosiaalisesti läsnäolevaksi.
    The tweets made me feel socially present.

16. Tunsin olevan yhteydessä muihin Twitter-käyttäjiin.
    I felt connected to other Twitter users.

17. Twiittejä oli mielestäni:
    I feel that the amount of tweets was…
    ❇ Answered on a scale from
    1 — "too low." to
    9 — "too high."

Lisätietoartikkelit / *Additional information*

(Questions 18 — 30 are omitted.)

Multitasking / *Multitasking*

Answers are given on the following scale:
(unless stated otherwise)

täysin eri mieltä *strongly disagree* **1 … 7** täysin samaa mieltä *strongly agree*

32. Olisin seurannut twiittejä myös ilman tablet-sovellusta.
I would have followed the tweets without a tablet application.

33. Olisin hakenut lisätietoa myös ilman tablet-sovellusta.
I would have searched for information without a tablet application.

34. Ponnistelin paljon pystäkseni seuraamaan sekä TV:tä että tablettia.
I struggled a lot with keeping up with both the TV ant the tablet.

35. Kuinka paljon koit katsoneesi tablet-laitetta ja TV-ruutua?
Which device did you follow more, the tablet or the TV screen?
❇ Answered on a scale from
1 — "I followed the tablet more" to
9 — "I followed the TV more"

36. Kuinka paljon koit katsoneesi twiittejä ja lisätietoartikkeleja?
How much did you follow the tweets and the information feed.
❇ Answered on a scale from
1 — "I focused more on the tweets." to
9 — "I watched the TV more on the articles."

37. Koin, että tabletti häiritsi TV-ohjelmaan keskittymistä.
I felt like the tablet distracted me from following the TV.

38. Koin, että TV häiritsi tablettiin keskittymistä.
I felt like the TV distracted me from following the tablet.

39. Koin, että päämedia on…
I consider _____ *as the main media.*
✓ Answered by choosing one of the following:
TV, tablet

Appendix D

Post-Test Questionnaire

Answers are given on the following scale:
(unless stated otherwise)

täysin eri mieltä *strongly disagree* **1** … **5** täysin samaa mieltä *strongly agree*

1. Kuvaile, miltä tablet-sovellus sinusta tuntui.
   Describe how you experienced the tablet application.
   ✎ Answered in writing.

2. Koin tablet-sovelluksen hyödylliseksi.
   The tablet application was useful.

3. Tablet-sovellusta käyttäessäni tunsin olevani yhteydessä muihin Twitter-käyttäjiin.
   While using the tablet application I felt connected with other Twitter users.

4. Katselukokemus oli mielyttävämpi, kun tabletilla näytettiin lisäsisältöä.
   The viewing experience was more pleasant when the tablet content was available.

5. Koin, että tabletti häiritsi ohjelmiin keskittymistä.
   I felt the tablet disturbed me from following the TV.

6. Koin, että TV häiritsi tablettiin keskittymistä.
   I felt the TV disturbed me from following the tablet.

7. Olisin valmis jakamaan TV:n katselutottumuksiani sosiaalisessa mediassa.
   I would be ready to share my TV viewing experience in social media.